10/52282037037

Rec'd PCT/PTO 2 4 JAN 2005

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

Organization
International Bureau

(43) International Publication Date
12 February 2004 (12.02.2004)

PCT

(10) International Publication Number
WO 2004/013327 A1

(51) International Patent Classification[7]: C12N 15/10,
C07K 14/00, C07H 21/00

(21) International Application Number:
PCT/EP2003/007765

(22) International Filing Date: 17 July 2003 (17.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02291871.8   24 July 2002 (24.07.2002)   EP

(71) Applicant *(for all designated States except US)*: LIBRA-GEN [FR/FR]; 11, Boulevard Einstein, F-69100 Villeur-banne (FR).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: NALIN, Renaud [FR/FR]; 54, Avenue du Général De Gaulle, F-69300 Caluire (FR). PUJIC, Petar [HR/FR]; 4, rue Chantemerle, F-69360 Serezin Du Rhone (FR). TUPHILE, Karine [FR/FR]; 181, rue de l'Oratoire, F-69300 Caluire (FR). GILLET, Benjamin [FR/FR]; 46, rue de la Thibaudière, F-69007 Lyon (FR).

(74) Agents: BECKER, Philippe et al.; Cabinet Becker et Associés, 35, rue des Mathurins, F-75008 Paris (FR).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/013327 A1

(54) Title: METHOD FOR THE EXPRESSION OF UNKNOWN ENVIRONMENTAL DNA INTO ADAPTED HOST CELLS

(57) Abstract: The present invention relates to methods for the identification or cloning of polynucleotides encoding a selected phenotype, particularly from environmental DNA. In a specific embodiment, the method comprises (i) cloning environmental DNA fragments into Ecoli cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

BEST AVAILABLE COPY

Method for the expression of unknown environmental DNA into adapted host cells

Introduction and Background

5   The present invention relates to methods and compositions for nucleic acid production, analysis and cloning. The present invention discloses tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify novel pathways, novel enzymes and novel metabolites of interest in various areas, including pharmaceutical, cosmetic,

10   agrochemical and/or food industry.

Drug discovery process is based on two main fields, namely combinatorial chemistry and natural products. Combinatorial chemistry has shown its ability to generate huge amounts of molecules, but with limited chemical diversity. At the opposite, natural

15   products have been the most predominant source of structural and molecular diversity. However, the exploitation of this diversity is strongly hampered by their limited access, complex identification and purification processes, as well as by their production.

Microorganisms are known to synthesize a large diversity of natural compounds which

20   are already widely used in therapeutic, agriculture, food and industrial areas. However, this promising approach to the identification of new natural compounds has always been considerably limited by the principal technological bolts of isolating and *in vitro* propagating the huge diversity of bacteria. Most microorganisms living in a natural, complex environment (soil, digestive tract, sea, etc...) have not been cultivated because

25   their optimal living conditions are either unknown or difficult to reproduce. Numbers of scientific publications relate this fact and it is now assumed that less than about 1% of the total bacterial diversity (when all environments are considered together) have been isolated and cultivated (Amann *et al*, 1995).

30   New approaches have been developed to try to overpass the critical step of isolation, and to access directly to the huge genetic potential established by the microbial adaptation

●

2

processes through their long evolution. These approaches are called "Metagenomic" because they address a plurality of genomes of a whole bacterial community, without any distinction (metagenome).

5   Metagenomics involve direct extraction of DNAs from environmental samples and their propagation and expression into a cultivated host cell, typically a bacteria. Metagenomic has been firstly developed for the identification of new bacterial phylum (Pace .1997). This use is based on the specific cloning of genes recognized for their interest as phylogenetic markers, such as 16S rDNA genes. Further developments of Metagenomics

10   relate to the detection and cloning of genes coding for proteins with environmental or industrial interest. These first two applications of metagenomic involve a first step of gene selection (generally using PCR) before cloning. In the case of protein production, the cloning vector used are preferentially also expression vectors, i.e., they contain regulatory sequences upstream of the cloning site causing expression of the cloned gene

15   in a given bacterial host strain.

More recent developments of metagenomic consider the total metagenome cloned without any selection and/or identification, to establish random "Metagenomic DNA libraries". This provides an access to the whole genetic potential of bacterial diversity without any "*a priori*" selection. Metagenomic DNA libraries are composed of hundreds

20   of thousands of clones which differ from each other by the environmental DNA fragments which have been cloned. In this respect, large DNA fragments have been cloned (more than 30 Kb), so as to (i) limit the number of clones which have to be analysed and (ii) to be able to recover whole biosynthetic pathways for the identification of new metabolites resulting from multi enzymatic synthesis. This last point is of

25   particular interest for bacterial metagenomic libraries since, most the biosynthetic pathways have been found to be naturally organised in a same cluster of DNA and even in the same operon in bacteria. Nevertheless, the heterologuous expression of a whole biosynthetic pathways (large DNA fragment) needs a much more improved system than a simple expression vector to have a full and stable expression.

30

Except for the identification and characterisation of bacterial community at the phylogenetic or diversity levels, metagenomic libraries produced in the prior art are gene expression libraries, i.e., the environmental DNA fragments are cloned downstream of a functional promoter, to allow their expression and analysis. In this regard, WO99/45154

5   and WO96/34112 relate to combinatorial gene expression libraries which comprise a pool of expression constructs where each expression construct contains DNA which is operably associated with one or more regulatory regions that drive expression of genes in an appropriate organisms. Furthermore, the expression constructs used in these methods have a very limited and invariable host range. Similarly, WO 01/40497 relates to the

10 construction and use of expression vectors which can be transferred in one chosen expression bacterial host of the *Streptomyces* genus. All these approaches are, however, very limited since they require the presence of expression signals and confer invariable or very limited host range capabilities. Furthermore, most (if not all) metagenomic DNA libraries have been established in *E. coli* which is the most efficient cloning system.

15 However, most environmental DNA are not expressed or functionally active in *E. coli*. In particular, functional analysis in *E. coli* of genes cloned from G+C rich organisms, such as Actinomyces, could be limited by the lack of adequate transcription and translation system. Also, posttranslational modification system in *E. coli* is not operative on heterologous proteins from Actinomicetes and some specific substrates for proteins

20 activity are not present in *E. coli*.

The stable maintenance of large foreign DNA fragments (> 10Kb) into a selected host cell is one of the key points for academic research or applied industrial purposes. Usually, the vector carrying the foreign DNA is maintained by cultivating the host cells

25 in a medium with a vector-specific selective pressure (resistance to an antibiotic for example). However, when large foreign DNA fragments are cloned and/or expressed, their propagation and/or expression require energy, which is not allocated for cell growth anymore. As a consequence of this new resource allocation (nutrients/energy), it is not unusual to have a genetic rearrangement of the foreign DNA (deletion, modification

30 etc...) as a recombinant cell reaction. This results in the modification of the foreign genetic information and in the loss of DNA functionality. This can be observed without

any loss of the selective pressure carried by the vector. As a result, the recombinant clone is no more exploitable for genetic or functional analysis.

5   Thus, the exploitation of the huge potential of metagenomics for the discovery of new natural compounds, pathways or genes cannot be achieved with currently existing methods. Alternative technologies and processes must be developed, to allow stable maintenance and propagation of large foreign DNAs into host cells for production of efficient libraries and functional screening in a large variety of host cell species, including *Bacillus* or *Streptomyces*, to take full account of the huge diversity of the

10   environmental DNAs.

Summary of the Invention

The present invention discloses improved tools and methods for the production and

15   analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify and produce novel pathways, novel enzymes and novel metabolites of interest.

More particularly, the invention now proposes to keep the advantage of high efficient

20   cloning in *E. coli* and to modify the properties of metagenomic libraries, to allow genetic and functional analyses of particular selected clones in any appropriate system, thereby making possible the stable maintenance and propagation, the analysis and/or the expression of the huge diversity of metagenomic libraries. According to the invention, polynucleotide libraries can be produced in any convenient cloning system, such as *E.*

25   *coli*, and then modified, depending on the desired selection or screening system, to adapt host range and/or properties of the library (or a portion thereof).

A particular object of this invention resides more specifically in a method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors

30   having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying

said selected cloning vectors to allow a transfer of said vectors into a selected host cell and integration of the polynucleotide contained in said vectors into the genome of the selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.

5

An other object of this invention is a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target

10   polynucleotide construct allowing (i) transfer of the environmental DNA into a selected host cell distinct from *E. coli*, (ii) integration of the environmental DNA into the genome of a selected host cell, and (iii) stable maintenance and propagation of the environmental DNA into the selected host cell.

15   A further object of this invention is a method of producing modified libraries of polynucleotides, the method comprising selecting a sub-population of clones in a first library, based on the presence or absence of a characteristic of interest, and modifying the properties of said selected clones to allow their functional analysis or expression.

20   The modification in the library or cloning vector is typically obtained by targeted insertion of a polynucleotide construct, preferably using transposable elements, either in vitro or in vivo.

      The integration into the genome of the selected host cell is typically obtained by site
25   specific integration or by homologous or heterologous DNA/DNA recombination.

      The invention is particularly suited for producing and analysing genetic diversity (metagenomic libraries), to identify new genes and isolate new metabolites, drugs, enzymes, antibiotics, etc.

30

6

Legend to the Figures

Figure 1 : Map of pP1 vector carrying transposable construct Tn<Apra> fig.1a. Excised transposable construct is 1132 bps in size. It contains two mosaic ends (ME) and a gene
5   conferring resistance to apramycine (Apra) fig.1b.

Figure 2 : pPL1 Vector (fig. 2a) that carries the conjugative transposable construct Tn<Apra-oriT> (fig. 2b). The nucleotide sequence of the transposable construct contains an origin of transfer (oriT) and a gene conferring resistance to apramycine (Apra).
10   Direction of DNA transfer at oriT is shown by an arrow.

Figure 3 : Transposable construct Tn<Apra-oriT-att-int> (fig. 3a) on vector pPAOI6 (fig. 3b). The transposable construct contains ΦC31 integrase gene and attachment DNA sequence for site specific integration, origin of transfer and gene for selection. The
15   orientation of genes and direction of DNA transfer are marked by arrows.

Figure 4 : This figure shows any target DNA suitable for insertion of transposable construct (fig. 4a). Insertion of conjugative and site specific integrative transposable construct Tn<Apra-oriT-int> is shown on fig. 4 b and c. Insertion of transposable
20   construct into selective gene marker carried on original vector is shown on fig. 4b. In this event, the cloned insert is intact and can be transferred to heterologous host. Insertion of transposable construct into cloned DNA insert results in gene inactivation (fig. 4c).

Figure 5 : Complete annotated DNA sequences of fosmid clones FS3-124 (a, SEQ ID
25   NO: 1) and FS3-135 (b, SEQ ID NO: 2).

Figure 6 : Morphological differences between *Streptomyces* transconjugant (assay). Conjugations have been performed with FS3-124 modified with transposable construct pPAOI6; (control) conjugation have been performed with pPAOI6.

30

Figure 7: Schematic Map of pPSB vector (top) and transposable elements (bottom). The transposable element has 720 bp DNA of the *amyE* gene from *B. subtilis* in addition to *att-int*-oriT-*Apra*.

5   Figure 8: Plasmid pPSBery (top) and transposable element (bottom) carrying selective marker *ery* AM, for resistance to erithromycine, and part of *amy E* gene for homologus recombination in *B. subtillis*.

Figure 9: Integrase Φ C31 was deleted from pPSBery plasmid. Resulting plasmid is
10   pPSBery-DI (top). Transposable element contains *Apra* and *ery* AM genes for selection, oriT origin of transfer and a part of *amyE* gene for integration in to *amyE* locus of *B. subtilis* chromosome (bottom).

Figure 10: Map of pTn5-7 AOI plasmid. Transposable element has ends of tn5 (ME) and
15   tn7 (T7 R, T7 L) transposons.

Detailed Description of the Invention

The invention provides novel strategies, methods and products for generating and
20   analysing combinatorial gene libraries. As indicated above, the invention discloses, particularly, methods of analysing libraries of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the methods comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a
25   transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

30   In a most preferred embodiment, the methods allow stable transfer and propagation of large environmental nucleic acids in a selected host following initial selection. Such

methods comprise (i) selecting cloning vectors in a library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and integration of the polynucleotide which they contain into a selected host cell genome, and (iii) analysing the polynucleotides
5   contained in said modified vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

Library of polynucleotides

10   The term "library of polynucleotides" designates a complex composition comprising a plurality of polynucleotides, of various origins and structure. Typically, the library comprises a plurality of unknown polynucleotides, i.e., of polynucleotides whose sequence and/or source and/or activity is not known or characterized. In addition to such unknown (or uncharacterized) polynucleotides, the library may further include known
15   sequences or polynucleotides. Typically, the library comprises more than 20 distinct polynucleotides, more preferably at least 50, typically at least 100, 500 or 1000. The complexity of the libraries may vary. In particular, libraries may contain more than 5000, 10 000 or 100 000 polynucleotides, of various origin, source, size, etc. Furthermore, the polynucleotides are generally cloned into cloning vectors, allowing their maintenance
20   and propagation in suitable host cells, typically in *E. coli*. The polynucleotides in the library may be in the form of a mixture or separated from each other, in all or in part. It should be understood that some or each polynucleotide in the library may be present in various copy numbers.

25   The polynucleotides in the libraries are more preferably obtained or cloned from complex sources of nucleic acids, most preferably from environmental samples. Such libraries are also termed "metagenomic libraries" since they contain nucleic acids derived from whole genomes of mixed populations of microorganisms.

30   The term environmental sample designates, broadly, any sample containing (a plurality of) uncharacterized (micro)organisms, particularly uncultivated (or non-cultivable)

microorganisms. The sample may be obtained or derived from specific organisms, natural environments or from artificial or specifically created environments (e.g., industrial effluents, etc). An uncultivated (or non-cultivable) microorganism is a microorganism that has not been purposely cultured and expanded in isolated form. The

5   sample may be obtained or derived from soil, water, mud, vegetal extract, wood, biological material, marine or estuarine sediment, industrial effluents, gas, mineral extracts, sand, natural excrements, meteorits etc. The sample may be collected from various regions or conditions, such as tropical regions, deserts, volcanic regions, forests, farms, industrial areas, household, etc.

10

Environmental samples usually contain various species of (uncharacterized, uncultivated) microorganisms, such as terrestrial microorganisms, marine microorganisms, salt water microorganisms, freshwater microorganisms, etc. Species of such environmental microorganisms include autotrophe or heterotrophe organisms,

15   eubacteria, archaebacteria, algae, protozoa, fungi, viruses, phages, parasites, etc. The microorganisms may include extremophile organisms, such as thermophiles, psychrophiles, psychrotophes, acidophiles, halophiles, etc. More specific examples of environmental bacteria includes actinomycetes, eubacteriaes and mycobacteriaes, examples of fungi include phycomycetes, ascomycetes and basidiomycetes, etc. Other

20   organisms include yeasts (saccharomyces, kluyveromyces, etc.) plant cells (algae, lichens, etc), corals, etc. for instance. The sample may comprise various species of such donor (uncultivated) microorganisms, as well as various amounts thereof. The environmental sample may contain, in addition, known and/or cultivable microorganisms (e.g., prokaryotic or eukaryotic), as well as nucleic acids and organic

25   materials. The sample may also contain different animal cells: mammalian cells; insect cells, etc (arising from larvae, feces, etc.).

It should be understood that the present invention is not limited to any specific type of sample or environmental microorganism, but can be used to produce diversity, create

30   nucleic acid libraries, etc., from any environmental sample comprising uncultivated microorganisms. The sample may be wet, soluble, dry, in the form of a suspension,

paste, powder, solid, etc. Preferably, the sample is dry or in solid or semi-solid state (e.g., paste, powder, mud, gel, etc.). The sample may be treated prior to nucleic acid extraction, for instance by washings, filtrating, centrifuging, diluting, drying, etc.

5   The term "environmental DNA" designates any DNA fragment or collection obtained from an environmental sample. Nucleic acids may be extracted / isolated from the sample according to various techniques, such as those described in WO01/81357, in WO01/40497, in Handelsman *et al.* (Chemistry & Biology 5(10), 1998, R245), Rondon *et al.* (Tibtech 17, 1999, 403 ; Applied and Environm. Microbiol. 66, 2000, 2541), Miller
10   et al (Applied and Environm. Microbiol. 65, 1999, 4715) or Frostegard *et al.* (Applied and Environm. Microbiol., 65, 1999, 5409).

In a particular embodiment of the above method, the library comprises a plurality of environmental DNA fragments. The library may also comprise other types of nucleic
15   acids, such as environmental RNAs, for instance.

The polynucleotides have a size which is typically comprised between 10 and 100 kb, more preferably between 20 and 80 kb, typically between 30 and 80 kb. Although not mandatory, it is preferred that the polynucleotide fragments in the library all have similar
20   size, to produce homogenous libraries.

Cloning Vectors

As indicated, the polynucleotides are contained or cloned into cloning vectors. These
25   vectors may be of various types, including plasmids, cosmids, fosmids, episomes, artificial chromosomes, phages, viral vectors, etc. In most preferred embodiments, the cloning vectors are selected from plasmids, cosmids, phages (e.g., P1-derivatives) and BACs, even more preferably from cosmids, P1 derivatives and BACs. By using cosmids or P1 derivatives, it is possible to generate homogenous libraries, since these vectors
30   essentially accommodate polynucleotides having a size of approximately 40 kb and 80 Kb, respectively. Furthermore, since the invention provides that the vectors are modified

after the initial cloning step, the cloning capacity of the vectors is maximized and inserts of as much as 40 and 80 kbs in length can be cloned into fosmids, BAC and P1 derivatives.

5   As indicated the cloning vector has a particular host range, i.e., the ability to replicate in a particular type of host cell. Typically, the host is a bacteria, more preferably an *E. coli* strain. Indeed, *E. coli* is so far the most convenient host cell for performing recombinant technologies. The advantage of the present invention is that the starting library can be produced in any suitable host system of choice, since the properties of the libraries will

10   be adapted later during the process.

Cloning vectors generally comprise the polynucleotide insert and genetic elements necessary and sufficient for maintenance into a competent host cell. They typically contain, in addition to the polynucleotide insert, an origin of replication functional in a

15   selected host cell as well as a marker gene for selection and screening. The cloning vector may comprise additional elements, such as promoter regions, for instance. Although cloning vectors may replicate in several different host cells, they are usually adapted to a particular host cell type and not suitable or efficient for replication or maintenance in other cell types.

20

In a preferred embodiment, the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids , BAC or P1 derived vectors. *E. coli* cloning vectors may carry an origin of replication derived from naturally-occurring plasmids, such as ColE1, pACYC and p15A, for instance. Many *E.coli* cloning vectors are commercially available and/or

25   can be constructed using available regulatory sequences.

Screening of the cloning vectors

In step i) of the method, a first selection or screen is performed on the polynucleotide

30   library. The screen is performed so as to identify or select clones having (or lacking) a particular, common characteristic. The selection may be carried out according to various

techniques, such as molecular screening, protein expression, functional screening, etc. A preferred selection is performed by molecular screening. Molecular screening designates any method of identification of molecular or structural characteristics in a polynucleotide sequence. This can be made by a variety of techniques which are known per se, such as

5   hybridisation, amplification, sequencing, etc. Preferably, molecular screening comprises the selection of clones in a library which contain, in their sequence, a particular sequence or region or motif, said sequence or region or motif being characteristic of a particular type of activity or gene (enzyme, biosynthetic pathways, etc.).

10   In a first variant, the selection is made by contacting the cloning vectors in the library with a particular nucleic acid probe (or set of probes) containing a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.). The cloning vectors in the library which hybridise to the probe (or set of probes) are then selected.

15

In a second variant, the selection is made by contacting the cloning vectors in the library with a particular pair of nucleic acid primers specific for a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.), and a PCR amplification reaction is performed. The cloning vectors in the library

20   which lead to a positive amplification product are then selected.

In this regard, the present application provides new primers designed in conserved motives of the β-keto acyl synthase gene, which are particularly useful for screening polynucleotides containing putative polyketide synthase (PKS) genes or domains. These

25   primers have the following degenerated sequence :
     Sense primer : 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3' (SEQ ID NO: 3)
     Antisense primer : 5'- GCBBSSRYYTCDATSGGRTCSCC -3' (SEQ ID NO: 4)
     wherein :
     R is A or G

30   S is G or C
     Y is C or T

K is G or T

D is A or G or T, and

B is C or G or T.

5   A particular object of this invention is a polynucleotide primer having one of the above sequences, typically a mixture of different polynucleotide primers having a sequence corresponding to one of the above degenerated sequences. A particular object of this invention also resides in a pair of primers each having one of the above sequences.

10   Once particular clones have been selected, the analysis of their polynucleotides needs to be confirmed and/or validated, and/or their polynucleotides can be used to study their function and/or produce novel compounds or metabolites.

The invention now enables such further analysis and uses, by allowing a modification of

15   the cloning vectors that is specific and adaptable by the skilled person, depending on the activity which is sought. In particular, it is possible to confer properties such as specific expression or a novel, specific host range to the selected cloning vectors, to assess their activity, as disclosed below.

20   <u>Modification of the cloning vectors</u>

After high efficiency cloning using most convenient cloning vectors such as BACs or cosmids propagated into *E. coli*, and after the identification, selection and/or characterisation of cloned DNA fragments, the invention now allows to modify

25   specifically the cloning vectors to transfer, integrate into the genome, maintain, express and/or over-express the selected polynucleotides into any selected host expression system, which is suitable to assess the selected activity or property. Such selected hosts may be native or heterologous host cells, and include, but are not limited to, for example *Streptomyces, Nocardia, Bacillus*, fungi, yeasts, etc.

30

The selected cloning vectors of the library may be modified according to various techniques. The modification is typically a genetic modification, comprising the introduction of particular genetic sequences into the structure of the cloning vector, in addition to or in replacement of sequences contained in said vector. It is highly preferred

5   to use specific or targeted (or oriented) techniques to improve the efficacy of the method. By "specific" is meant that the modification occurs at a pre-determined location in the cloning vector, through site-specific mechanisms. By "targeted" is meant that the modification occurs in a controlled way, so as not to alter the polynucleotide insert contained in the vector in a non-desirable way.

10

In a preferred embodiment, the selected vectors are modified by insertion, into the vector, of a target polynucleotide construct which contains genetic elements conferring the selected property(ies) to the cloning vector.

15   The target polynucleotide construct typically comprises the genetic elements necessary to transfer, propagate, integrate into the genome, maintain, express or overexpress the cloned polynucleotide into a chosen (bacterial) host expression system. Said genetic elements may include particular origin(s) of replication, particular origin(s) of transfer, particular integrase(s), transcriptional promoter(s) or silencer(s), either alone or in

20   combination(s).

In a first, preferred variant, the target polynucleotide construct comprises a genetic element allowing transfer of the vector into a selected host cell.

25   Natural DNA transfer mechanisms between donor and recipient strains is known under the term conjugation or conjugative transfer. Conjugative transfer can occur between different strains of the same species as well as between strains of different species. Many naturally occurring plasmids carry so called *tra* genes, which are involved in and mediate conjugative transfer. The DNA transfer starts at specific DNA structures, known as an

30   origin of transfer or "ori T". The presence of such an oriT in a vector allows said vector to be transferred into a desired host cell.

In a particular, preferred embodiment, the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.

The structure of various oriT has been reported in the art (Guiney *et al.* 1983; Zechner *et*

5   *al.* 2000). In a specific embodiment, the origin of transfer is selected (or derived) from RP4, pTiC58, F, RSF1010, ColE1 and R6K(α).

A specific example of an oriT which can be used in the present invention derives from plasmid RP4 and has or comprises all or a functional part of the following sequence

10 (SEQ ID NO: 5) :

```
gatctGTGATGTACTTCACCAGCTCCGCGAAGTCGCTCTTCTTGATTGGAGCGCATGGG
GACGTGCTTGGCAATCACGCGCACCCCCCGGCCGTTTTAGCGGCTAAAAAAGTCAT
GGCTCTGCCCTCGGGCGGACCACGCCCATCATGACCTTGCCAAGCTCGTCCTGCTTC
TCTTCGATCTTCGCCAGCAGGGCGAGGATCGTGGCATCACCGAACCGCCGTGCG
CGGGTCGTCGGTGAGCCAGAGTTTCAGCAGGCCGCCCAGGCGGCCCAGGTCGCCAT
TGATGCGGGCCAGCTCGCGGACGTGCTCATAGTCCACGACGCCCGTGATTTTGTAGC
CCTGGCCGACGGCCAGCAGGTAGGCCGACAGGCTCATGCCGGCCGCCGCCGCCTTT
TCCTCAATCGCTCTTCGTTCGTCTGGAAGGCAGTACACCTTGATAGGTGGGCTGCCC
TTCCTGGTTGGCTTGGTTTCATCAGCCATCCGCTTGCCCTCATCTGTTACGCCGGCGG
TAGCCGGCCAGCCTCGCAGAGCAGGATTCCCGTTGAGCACCGCCAGGTGCGAATAA
GGGACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCTATCCTGCC
CGGCTGACGCCGTTGGATACACCAAGGAAAGTCTACACGAACCCTTTGGCAAAATC
CTGTATATCGTGCGAAAAAGGATGGATATACCGAAAAAATCGCTATAATGACCCCG
AAGCAGGGTTATGCAGCGGAAAAGATCCGTCGGATCT
```

The term "functional part" designates any fragment or variants of the above sequence which retain the capacity to cause conjugative transfer. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants

30 may include one or several mutations, substitutions, deletions or additions of one or several bases.

In an other particular variant, the target polynucleotide construct comprises a genetic element allowing integration of the vector (or of the polynucleotide contained therein)

35 into the genome of the selected host cell.

A donor DNA is permanently or stably maintained and expressed in a selected recipient cell if it is integrated into the recipient cell's genome or if it contains elements that allow autonomous replication in said cell. In a most preferred embodiment, the vector is modified to allow transfer and integration of the polynucleotide into the host cell
5   genome.

Integration is a preferred way of ensuring stable expression. Integration can be obtained by physical recombination. Recombination can be homologous, e.g., between two homologous DNA sequences, or illegitimate, where recombination occurs between two
10   non-homologous DNAs. As a particular example, integration of donor DNA into the chromosome of the recipient can be mediated by host recombination repair system or by site-specific recombination. Another well-studied process that can transfer and integrate genes is transduction by bacterial viruses, such as λ and φC31. In a phage–infected bacterial cell, fragments of the host DNA are occasionally packaged into phage particles
15   and can then be transferred to a recipient cell. Integration into the recipient cell's genome is caused by an integrase.

In a specific embodiment, the target polynucleotide construct comprises a nucleic acid encoding an integrase functional in the selected host cell. More preferably, the integrase
20   is selected from λ and φC31 integrases. In a specific embodiment, the polynucleotide construct comprises a nucleic acid encoding an integrase having or comprising all or a functional part of the following sequence of the φC31 integrase (SEQ ID NO: 6) :

```
     AGATCTCCCGTACTGACGGACACACCGAAGCCCCGGCGGCAACCCTCAGCGGATGC
25   CCCGGGGCTTCACGTTTTCCCAGGTCAGAAGCGGTTTTCGGGAGTAGTGCCCCAACT
     GGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGGGTCGCCGACATGACACAAG
     GGGTTGTGACCGGGGTGGACACGTACGCGGGTGCTTACGACCGTCAGTCGCGCGAG
     CGCGAGAATTCGAGCGCAGCAAGCCCAGCGACACAGCGTAGCGCCAACGAAGACA
     AGGCGGCCGACCTTCAGCGCGAAGTCGAGCGCGACGGGGGCCGGTTCAGGTTCGTC
30   GGGCATTTCAGCGAAGCGCCGGGCACGTCGGCGTTCGGGACGGCGGAGCGCCCGGA
     GTTCGAACGCATCCTGAACGAATGCCGCGCCGGGCGGCTCAACATGATCATTGTCT
     ATGACGTGTCGCGCTTCTCGCGCCTGAAGGTCATGGACGCGATTCCGATTGTCTCGG
     AATTGCTCGCCCTGGGCGTGACGATTGTTTCCACTCAGGAAGGCGTCTTCCGGCAGG
     GAAACGTCATGGACCTGATTCACCTGATTATGCGGCTCGACGCGTCGCACAAAGAA
35   TCTTCGCTGAAGTCGGCGAAGATTCTCGACACGAAGAACCTTCAGCGCGAATTGGG
     CGGGTACGTCGGCGGGAAGGCGCCTTACGGCTTCGAGCTTGTTTCGGAGACGAAGG
```

```
AGATCACGCGCAACGGCCGAATGGTCAATGTCGTCATCAACAAGCTTGCGCACTCG
ACCACTCCCCTTACCGGACCCTTCGAGTTCGAGCCCGACGTAATCCGGTGGTGGTGG
CGTGAGATCAAGACGCACAAACACCTTCCCTTCAAGCCGGGCAGTCAAGCCGCCAT
TCACCCGGGCAGCATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCCGA
CCCGGGGCGAGACGATTGGGAAGAAGACCGCTTCAAGCGCCTGGGACCCGGCAACC
GTTATGCGAATCCTTCGGGACCCGCGTATTGCGGGCTTCGCCGCTGAGGTGATCTAC
AAGAAGAAGCCGGACGGCACGCCGACCACGAAGATTGAGGGTTACCGCATTCAGCG
CGACCCGATCACGCTCCGGCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCCG
CTGAGTGGTATGAGCTTCAGGCGTGGTTGGACGGCAGGGGGCGCGGCAAGGGGCTT
TCCCGGGGGCAAGCCATTCTGTCCGCCATGGACAAGCTGTACTGCGAGTGTGGCGC
CGTCATGACTTCGAAGCGCGGGGAAGAATCGATCAAGGACTCTTACCGCTGCCGTC
GCCGGAAGGTGGTCGACCCGTCCGCACCTGGGCAGCACGAAGGCACGTGCAACGTC
AGCATGGCGGCACTCGACAAGTTCGTTGCGGAACGCATCTTCAACAAGATCAGGCA
CGCCGAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGGAAGCCGCCCGACGCTTCG
GCAAGCTCACTGAGGCGCCTGAGAAGAGCGGCGAACGGGCGAACCTTGTTGCGGAG
CGCGCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGGCAGGCGC
GTACGACGGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGGCAGCGCTGACGC
TCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAACTTGAAGCCGCCGAAGCCCCG
AAGCTTCCCCTTGACCAATGGTTCCCCGAAGACGCCGACGCTGACCCGACCGGCCCT
AAGTCGTGGTGGGGGCGCGCGTCAGTAGACGACAAGCGCGTGTTCGTCGGGCTCTT
CGTAGACAAGATCGTTGTCACGAAGTCGACTACGGGCAGGGGGCAGGGAACGCCCA
TCGAGAAGCGCGCTTCGATCACGTGGGCGAAGCCGCCGACCGACGACGACGAAGAC
GACGCCCAGGACGGCACGGAAGACGTAGCGGCGTAGCGAGACACCCG
```

The term "functional part" designates any fragment or variants of the above sequence which retain the capacity to cause integration. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants may include one or several mutations, substitutions, deletions or additions of one or several bases.

In a more preferred variant, the target polynucleotide construct comprises genetic elements allowing transfer of the cloning vector into the selected host cell and integration of the cloning vector or a portion thereof into the genome of the selected host cell. Most preferred polynucleotide constructs comprises an oriT and a nucleic acid encoding an integrase.

In an other variant, the target polynucleotide construct comprises an origin of replication specific for or functional in the selected host cell. The origin of replication may be selected (or derived), for instance, from pAMβ1, pSa, 2μm circle, pSam2, pSG1, pIJ101, SCP2, pA387 and artificial chromosomes.

In an other variant, the target polynucleotide construct comprises a transcriptional promoter functional in the selected host cell. As indicated above, in a particular variant, the invention allows to modify the cloning vector to enable expression or over-expression of the cloned polynucleotides in the selected host. The expression of genes is

5 driven mainly by transcriptional promoters, which initiate gene transcription. The type of promoter to be used in the present invention can be selected by the skilled person, depending on the selected host cell and type of expression needed. Promoters may be ubiquitous or cell-specific, regulated or constitutive, weak or strong. They may be of various origins, including promoters isolated from viruses, phages, plant cells, bacterial

10 genes, mammalian genes, etc., or they may be artificial or chimeric. Typical examples of promoters include T7, T4, LacZ, trp, ara, SV40, tac, λPL, GAL, AOX, hsp-70, etc.

The target polynucleotide construct is typically a DNA molecule, although RNAs may also be used as starting material. It is typically a double-stranded DNA. The target

15 polynucleotide construct may be produced by conventional recombinant DNA techniques, including DNA synthesis, cloning, ligation, restriction digestion, etc. and a combination thereof.

The target polynucleotide construct is preferably engineered so as to be inserted in a

20 region of the vector distinct from the polynucleotide. Indeed, it is important that the integrity of the polynucleotides is preserved. Directed insertion may be accomplished in a variety of ways, including site-specific insertion using particular enzymatic systems (Cre/Lox, FLP, etc.), homologous recombination with particular target sequences present in the vector, or by the use of transposons or transposable elements and appropriate

25 selection means.

In a particular, preferred embodiment, the target polynucleotide construct is contained in or comprises a transposable nucleic acid construct. Indeed, in a preferred variant, the methods of the present invention use transposable elements to alter the properties of the

30 cloning vectors, and allow their transfer, maintenance, expression or over-expression in a selected host cell.

Transposable nucleic acid constructs are derived from transposons, which are genetic elements capable of moving from one genetic loci to another. Two main classes of transposons have been identified in bacteria. The most simple transposons comprise an

5   insertion sequence that carries only elements of transposition. These elements are two inverted DNA repeats and a gene that codes for a protein called transposase. The transposase catalyses the excision and integration of the transposon. It has been shown that the excision and integration reaction can be catalysed in trans by a transposase, which can be provided in vivo or in vitro in purified form or expressed from a different

10   construct. More complex transposons carry more insertion sequences and additional genes that are not involved in transposition.

Transposable nucleic acid constructs of this invention thus typically comprise, flanked by two inverted repeats, the target polynucleotide construct and, more preferably, a

15   marker gene. In the presence of a transposase, these transposable nucleic acid constructs can integrate into a cloning vector in vivo or in vitro, thereby providing for targeted polynucleotide insertion. Alternatively, such nucleic acid constructs can be used for targeted integration, in the absence of a transposase, in particular strains such as hypermutator strains. Such transposable nucleic acid constructs also represent a

20   particular object of the present application. In this regard, in a more preferred embodiment, the invention also relates to a transposable nucleic acid construct, wherein said construct comprises an origin of transfer flanked by two inverted repeats. Specific examples of such construct are transposons pPL1 and pPAOI6, as disclosed in the experimental section. The transposable nucleic acid construct may further comprise an

25   integrase gene and/or a marker gene.

The inverted repeat nucleic acid sequences may be derived from the sequence of various transposons, or artificially created. In particular, transposable elements can be generated using inverted repeats obtained from transposons or transposable elements such as Tn5,

30   Tn21, miniTn5, T7, T10, Tn917, miniTn400, etc. Preferably, the sequences derive from

20

transposon Tn5. In a specific embodiment, they comprise all or a functional part of the following sequences :

- left arm of pPAOI6 transposon (SEQ ID NO: 7)
5   CTGTCTCTTATACACATCTCAACCATCATCGATGAATTTTCTCGGGTGTTCTCGCATA
    TTGGCTCGAATTCGAGCTCGGTACCC

- right arm of transposon pPAOI6 (SEQ ID NO: 8)
    GATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGCCAACGACTACGCACTAGCC
10 AACAAGAGCTTCAGGGTTGAGATGTGTATAAGAGACAG

The marker gene may be any nucleic acid encoding a molecule whose presence in a cell can be detected or visualized. Typical marker genes encode proteins conferring resistance to antibiotics, such as apramycine, chloramphenicol, ampiciline, kanamycine,
15 spectinomycine, thiostrepton, etc. Other types of markers confer auxotrophy or produce a label (e.g., galactosidase, GFP, luciferase, etc).

In a specific embodiment, the cloning vector in the library comprises a first marker gene and the modification step ii) comprises:
20 . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and
    . selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.
25

The double selection ensures that the target polynucleotide construct has been inserted at a site within the marker gene present in the cloning vector, i.e., outside of the polynucleotide insert.

30 It should be understood that the modification may be accomplished in various other ways, particularly by incorporating a sequence coding for the transposase directly into the transposable element or into another expression unit. The presence and expression of the transposase can be regulated by inductive promoter or termosensitive replicative units. Also, transposition can be carried out by *in vitro* process.

Analysis of the polynucleotides

In step (iii) of the process, the polynucleotides may be analysed by various methods,
5   including by genetic, biochemical, chemical or phenotypical approaches, which are well
known per se in the art. Analysis occurs upon transfer and, optionally, expression of the
polynucleotides into the selected host cell.

In this regard, the modified cloning vectors can be transferred into the selected host cell
10   by a variety of techniques known in the art, including by transformation, electroporation,
transfection, protoplast fusion, conjugative transfer, etc. In a preferred embodiment, the
target polynucleotide construct comprises an oriT and the modified vectors are
transferred into the selected host cells by conjugative transfer. In this embodiment, the
cloning vector and the selected host cells are co-cultivated and the recombinant host cells
15   are selected and isolated.

The selected host cell may be any type of cell or microorganism, including, without
limitation, *Steptomyces, E. coli, Salmonella, Bacillus, Yeast, fungi*, etc.

20   One of the objectives of the invention is to be able to analyse environmental DNAs of
unknown cellular origin into different host expression systems. In order to analyse the
potentiality of the DNAs at the transcription and/or translation levels and to have much
more probabilities to have a DNA expression, it is important to have the possibility to
test different host expression systems.

25

Insertion of foreign DNA into host expression systems like *Streptomyces* can produce,
for instance, an increase in doubling time, morphological modifications, pigments
production, etc., which can be related either directly to the expression of foreign DNA or
by combinatorial biology of the foreign DNA and the biology of the expression host
30   systems. The new phenotypes can be analysed by all techniques known in the art such
genetic, biochemical, chemical, phenotypic approaches etc.

In a preferred, specific embodiment, the invention relates to a method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E. coli* cloning vectors to
5   produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments
10   contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

By applying the above method, new polynucleotide sequences have been identified, cloned and characterized, which produce new phenotypes in bacteria. These
15   polynucleotides contain the sequence of PKS genes and other genes that encode polypeptides involved in biosynthetic pathways. The sequence of these polynucleotides is provided in Figure 5. The invention also relates to any polynucleotide sequence comprising all or part of these sequences (i.e., SEQ ID NOs: 1 or 2), their complementary strand, or a functional variant thereof. A part of the above sequences
20   includes, preferably, at least 20 consecutive bases, more preferably at least 50 consecutive bases thereof, even more preferably a coding sequence (e.g., an CDS). In this respect, SEQ ID NOs: 1 and 2 comprise several novel open reading frames encoding novel polypeptides involved in biosynthetic pathways. These coding sequences are identified in Figures 5a and 5b.

25

In a specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides (CDS) 76 - 1134 ; 1096 - 2430 ; 1178 - 1624 ; 2506 - 3567 ; 2906 - 4222 ; 4092 - 5321 ; 6337 - 8502 ; 8181 - 9530 ; 9531 - 10721 ; 10504 - 11274 ; 12874 - 13689 ; 14195 - 15976 ; 15427 - 16512 ; 15579 - 16253 ; 16505
30   - 17656 ; 17657 - 18697 ; 18615 - 19304 ; 19301 - 20596 ; 20535 - 21476 ; 22025 -

22951 ; 23155 - 26523 ; 26409 - 34433 ; 34418 - 37500 and 35359 - 37500 of SEQ ID NO: 1 (figure 5a) or a complementary strand thereof.

5   In an other specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides 3 - 914, 924 - 2168 ; 2207 - 3190 ; 3373 - 4455 ; 4546 - 4959 ; 5176 - 6192 ; 6331 - 14043 ; 14275 - 15408 ; 15436 - 16245 ; 16287 - 17384 ; 17427 - 18158 ; 18248 - 18847 ; 18952 - 20346 ; 20442 - 21167 ; 21164 - 24301 ; 24351 - 27023 ; 27806 - 29686 ; 29535 - 30872 ; 30848 - 32647 ; 32574 - 35555 ; 35533 - 36598 and 36516 - 37400 of SEQ ID NO: 2 (figure 5b) or a 10   complementary strand thereof.

Variants of these sequences include any naturally-occurring variant comprising or or several nucleotide substitutions ; sequences variants resulting from the degeneracy of the genetic code, as well as synthetic variants coding for functional polypeptides. Variants 15   include any sequence that hybridise under high stringent conditions, as disclosed for instance in Sambrook et al., to any of the above sequences, and encode a functional polypeptide. The invention also include any nucleic acid molecule encoding a polypeptide comprising all or a fragment of an amino acid sequence encoded by a polynucleotide as disclosed above. Preferably, the fragment comprises at least 10 20   consecutive amino acid residues, more preferably at least 20, even more preferably at least 30.

The invention also relates to any vector comprising these polynucleotide sequences. These sequences may be DNA or RNA, preferably DNA, even more preferably double-25   stranded DNA. The invention also relates to a polypeptide encoded by a polynucleotide sequence as defined above. The invention also relates to a method of producing such polypeptides by recombinant techniques, comprising expressing a polynucleotide as defined above in any suitable host cell and recovering the encoded polypeptide. The invention also relates to a recombinant host cell comprising a polynucleotide or a vector 30   as defined above.

An other object of this invention resides in a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a
5   target polynucleotide construct allowing transfer and integration of the environmental DNA into the genome of a selected host cell distinct from *E. coli*.

The sub-DNA-libraries should have either desired genetic characteristics based on high or low GC content, DNA encoded for a desired enzymatic activity, part or full
10   biosynthetic pathways for metabolites etc., or specific origin such as soil fractions, animal organs, sub fraction of a microorganism community etc. The invention also allows to produce conjugative vector with desired characteristics in accordance with the characteristics of the pre-identified sub-DNA-libraries and functional analysis of mutants in heterologous hosts. It can also be used, without limitation, for the production of
15   mutants by mutagenesis, for DNA sequencing, genes or biosynthetic pathways knock-out by insertion or to confer transfer capabilities for expression, co-expression, over-expression or modification of biosynthetic pathways.

Further aspects and advantages of the invention will be disclosed in the following
20   examples, which should be regarded as illustrative and not limiting the scope of this application.

Experimental Section

25   **A - From *E.coli* to *Streptomyces***

In this work, we constructed a fosmid library in *E. coli* from total DNA prepared directly from soil. The library has been screened for presence of biosynthetic pathways. We developed genetic tools for functional genomics that allow gene identification,
30   inactivation and horizontal gene transfer from *E. coli* to *Streptomyces*.

25

The cloning vectors in the library contain ColE1 replicon for propagation in *E. coli*. Transposable elements based on Tn 5 transposon were produced and used for *in vitro* modification of selected cloning vectors. Integrated transposable elements contain gene for resistance to apramicyne. Conjugative derivatives were constructed by incorporating

5   origin of transfer from plasmid RP4. A conjugative and site specific integrative transposon was also constructed comprising the integrase gene from φ C31 phage, including attP attachement site. Conjugal transfer was demonstrated from an appropriate *E. coli* donor cell to another *E. coli* or *Streptomyces lividans* recipient cell. Constructed transposon was tested for inactivation of the genes cloned into fosmids. Obtained

10   mutants can be used for direct sequencing by adequate primers and transferred by conjugation into *Streptomyces lividans*. Transposable elements thus represent very useful tools for functional analysis of a large DNA libraries cloned into BAC, PAC fosmids or other cloning vectors in which cloned inserts must be transferred into heterologous host.

Materials and methods

15   Bacterial strains, plasmids and growth conditions.

E. coli DH10B (F- mcrA delta(mrr-hsdRMS-mcrBC) phi80dlacZ deltaM15 delta lacX74 deoR recA1 endA1 araD139 delta (ara, leu)7697 galU galK lambda- rpsL nupG), strain (Epicentre) was used for fosmid and plasmid transformation and DNA amplification.

Unless specifically described, all DNA manipulations were performed according to

20   Sambrook, J., *et al*, (1989).

Soil DNA extraction and DNA libraries construction.

Total bacterial community DNA was extracted and large DNA libraries have been constructed into fosmids according to the method described in WO 01/81357.

Fosmid DNA extraction and Purification

25

Fosmids DNA containing soil librairy were extracted from pools of 96 clones. Culture of recombinant clones were performed in Deep-Well 96 and 48, respectively in 1 ml and 2 ml of LB media containing 12,5 µg ml^{-1} of chloramphenicol. Cultures were grown at

● ●

26

37°C with shaking at 250 RPM during 22 hours. DNA extraction was done by using the Nucleobond PC100 extraction kit (Macherey Nagel).

PCR screening for the detection of PKS genes

5   Primers design

Degenerate PCR primers sets were designed to specifically amplify PKS nucleic acids sequences. Multiple sequence alignment of PKS domains revealed highly conserved motives, in particular in the β-keto acyl synthase domain. Primers Lib1F and Lib2R were designed in conserved motives of the β-keto acyl synthase gene. Lib1F (sense

10   primer, 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3') and Lib2R (antisense primer, 5'- GCBBSSRYYTCDATSGGRTCSCC -3') were deduced from β-keto acyl synthase peptide sequences GP(AS)(LV)(AST)(IV)DTAC and GDPIE(TVA)(RAQ)A, respectively. The specific fragment amplified with Lib1F / Lib2R was approximately about 465 bp (corresponding to 155 amino acids). Specificity and efficiency of the PCR

15   systems were validated by testing on positive DNA controls (i.e. genomic DNA from type I PKS producing strain such as *Bacillus subtilis, Streptomyces lividans, Streptomyces ambofaciens* and *Ralstonia solanocearum*) and negative DNA controls (genomic DNA from strains which are known to do not contain PKS genes). Furthermore, DNA extracted from soil samples were tested to calibrate PCR techniques.

20

          PCR conditions

PCR conditions were optimised, in particular for concentrations of DMSO, $MgCl_2$, Primers and DNA template quantities. For PCR using microorganism genomic DNA and soil DNA as template (50 to 200 ng), the PCR mix (50μl) contained 250μM of dNTP, 5

25   mM $MgCl_2$ final, 2,5% DMSO, 1X PCR buffer, 0,75 μM of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For PCR using fosmid pooled DNA as template (100 to 500 ng), the PCR mix (50μl) contained 250μM of dNTP, 5 mM $MgCl_2$ final, 5% DMSO, 1X PCR buffer, 0,75 μM of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For identification of positives

30   clones in 96 microtiter plates, 25 μl of each bacteria culture were used as template and PCR conditions were the same as above. Thermocycling program was : a denaturation

step at 96°C for 5 minutes; then 1 minute at 96°C, 65°C for 1 minute, 72°C for 1 minute. The first 7 cycles, the annealing temperature was lowered 1°C per cycle until 58°C was reached. A subsequent 40 cycles were carried out with the annealing temperature at 58°C. A final extension step was at 72°C for 7 minutes. For identification of positives

5   clones in 96 microtiter plates, the first denaturation step of 96°C was during 8 minutes. The other steps were the same as described above. PCR reactions were performed with a PTC 200 thermocycler (MJ Research).

PCR products analysis

10   PCR products of about 465 bp were purified on agarose gel with gel extraction Kit (Qiagen) according to the manufacturer recommendations. First approach consisted in subcloning PCR products using the Topo PCR II kit (Invitrogen). Recombinant Plasmids were extracted using QIAprep plasmid extraction Kit (Qiagen) and sequenced with Forward and Reverse M13 primers with CEQ 2000 automated sequencer (Beckman

15   Coulter). Second approach consisted in direct sequencing of PCR products. Sequencing data were compared with nucleic and proteic genbank database using BLAST program.

Sequencing of the identified fosmid insert DNA and sequence analysis

Fosmids inserts were sequenced using either a transposon-mediated and by shotgun

20   subcloning approach. Transposition was realized by using (Transposition Kit) commercialized by Epicentre according to the manufacturer. For shotgun subcloning, transformants were grown for 16 hours at 37°C. Fosmid extraction was done by using the Nucleobond PC100 extraction kit (Macherey Nagel). DNA was partially restricted with *Sau3A* and sized on standard gel elecrtphoresis for fragments ranging from 1 to 3

25   Kbs and cloned into Bluescript vector according to Sambrook *et al.* (1989).

Sequence analysis was performed with the identification of ORFs by using Frameplot of the GC3. Each identified ORF was compared to gene databases by using BLAST program. PKS domains were determined by aligning obtained sequence versus already described PKS domains from domain databases.

Sequencing

Sequencing reactions were performed with 1 μg of DNA and 3.6 pmol of primer, using CEQ 2000 Dye Terminator Cycle Sequencing kit (Beckman Coulter) under conditions proposed by supplier. Ten μL of reaction products were precipitated using 4 μL of
5   solution containing 1.5 M NaOAc, 50 mM EDTA and 60 μL cold 95% ethanol/dH$_2$O from –20 °C. The pellet was washed 2 times with 200 μL 70% ethanol/dH$_2$0, vacuum dried and dissolved in 40 μL sample loading solution (supplied in kit). Sequencing reactions were run on an CEQ 2000 sequencer (Beckman Coulter).

10   Plasmids construction and validation

Plasmid pP1 was constructed as follows. A 941 bp DNA fragment containing native promoter region and AA(3)IV gene was amplified by polymerase chain reaction (PCR) using primers AmF (d-CCCTAAGATCTGGTTCATGTGCAGCTCCATC, SEQ ID NO: 9) and AmR (d-TAGTACCCGGGGATCCAACGTCATCTCGTTCTCC, SEQ ID
15   NO: 10). One hundred microlitre reaction were performed containing 0.1μM each of primers, 1 X Vent DNA polymerase buffer (NEB), 0.2 μM of each deoxyribonucleoside triphosphate (dNTP), 50 ng of the DNA template and 2U of Vent DNA polymerase (NEB). PCR mixture was heated for 4 min at 94 °C in a PTC-200 thermocycler (Peltier) and cycled 25 x at 94 °C for 60 sec, 59 °C for 30 sec and at 72 °C for 70 sec. The final
20   extension was performed at 72 °C for 7 min.

PCR product was purified using GFX DNA purification kit (Amersham), then digested by *Bgl* II and *Sma* I restriction enzymes. A 941 bp BglII/SmaI fragment was inserted into the *Bam* HI, *Sma* I sites of pMOD plasmid (EPICENTRE). DH10B *E. coli* was transformed with pP1 and subjected to apramicyne selection on LB agar plates. Six
25   colonies surviving on apramicyne selection were grown in liquid LB media and final pPL1 candidates were thoroughly checked via PCR and restriction mapping.

To construct conjugative plasmid pPL1, 750 bp oriT DNA region from plasmid RP4 were   amplified   via   PCR   using   primers   oriTF   (d-
30   GCGGTAGATCTGTGATGTACTTCACCAGCTCC, SEQ ID NO: 11) and oriTR

29

(TAGTACCCGGGGATCCGACGGATCTTTTCCGCTGCAT, SEQ ID NO: 12). PCR conditions were as above. Amplified DNA was digested using *Bgl* II and *Sma*I restriction enzymes. *BglII/Sma*I DNA fragment was subjected to purifiation after gel electophoresis on 0.7% agarose. Purified fragment was ligated into pP1 plasmid digested by *Bam* HI

5   and *Sma* I restriction enzymes.

φ C31 integrase gene and attachment site (attP) was amplified via PCR using primers Fint (d-AACAAAGATCTCCCGTACTGACGGACACACCG, SEQ ID NO: 13) and PJ (d-CGGGTGTCTCGCATCGCCGCT, SEQ ID NO: 14). Amplified DNA fragment was

10   purified by GFX kit (Amersham) and phosphorilated using T4 polynucleotide kinase (NEB) under conditions recommended by the enzyme manufacturer. Phosphorilated DNA fragment was cloned in to pPL1 vector opened with SmaI restriction enzyme (NEB) and dephosphorilated by calf alkaline phosphatase (NEB). DH10B *E coli* was transformed with ligation mixture using Bio Rad Pulsing apparatus and protocols

15   provided by Bio-Rad. Twelve transformants were analyzed by PCR for the presence of integrase gene. Orientation of integrase gene was verified by restriction analysis using *Bgl* II and *Eco*RI restriction enzymes. Resulting plasmid was named pPAOI6.

To construct pPAOI6-A plasmid, pPAOI6 plasmid was digested with *Eco*RI and *Bgl*II restriction enzymes followed bydigestion by Bean mung nuclease (NEB). Linearised

20   plasmid was self ligated and transformed in to DH10B cells.

Plasmid preparation of fosmid DNA

Fosmid and BAC DNA for sequencing was prepared by using the Nucleobond AX kit (Macherey-Nagel), following protocol for BACs, Cosmid as specified by manufacturer.

Mutagenesis

25   Transposon Tn-pPAOI6 was prepared by digestion of pPAOI6 plasmid using *Pvu*II restriction enzymes, followed separation on agarose gel and purification of fragment containing transposon from gel using Qiagen kit. The same molar ratio of transposon and corresponding fosmid was used for mutagenesis in vitro using Tn5 transposase (Epicentre) and conditions specified by manufacturer. We transformed aliquot of the

30   transposed mixture by electroporation into competent DH10B *E. coli* strain.

Conjugation *E. coli- Streptomyces lividans TK24*

Conjugation experiments were done using 6 x 10^6 *E coli* S17.1 cells containing conjugative plasmids or fosmids. The *E coli* cells were grown in LB media with adequat antibiotic. The cells were collected by centrifugation, washed two times using same

5 volume of LB media and concentrated to 10^8 cells/ml and overlaid on LB plates containing 2x 10^6 pregeminated *Streptomyces lividans* TK24 spores. The cells mixture were grown over night at 30^0 C and *E. coli* cells were washed three times using 2 ml of LB media. The plates were overlayed using top agar containing NAL (nalidixic acid) and the appropriated antibiotic. Plates were incubated for 4 days at 30°C and transformant

10 streptomyces colonies were isolated on HT medium (Pridham *et al.* 1957) containing NAL and the same appropriated antibiotic.

Results

Construction of a Transposon Tn <Apra>.

15 *E. coli* aminoglycoside -(3)- acetyl transferase IV gene (aa(3)IV) was amplified by PCR and cloned in pMOD vector (Epicentre) . Advantage of this selective marker allow positive selection in *E. coli* and in *Streptomyces lividans*. Transposon can be used for insertional inactivation in vitro using purified transposase Tn5 (Epicentre). The structure of the pP1 constructed vector and transposon was shown on Figure 1a,1b.

20 Construction of a conjugative plasmid-transposon.

Conjugative vector, transposon was constructed by cloning origin of transfer from plasmid RP4 into pP1 vector producing pPL1 vector (Figure 2a, 2b). The origin of transfer was cloned in such orientation that the selective aa(3)IV gene is the last transferred during conjugation. PPL1 vector was introduced in to specific *E. coli* S17.1

25 strain that carry RP4 plasmid integrated in to chromosome. In conjugation experiment between donor strain S17.1 carrying pPL1 plasmid and DH10 *E coli* receptor we obtained DH10B strain carrying pPL1 plasmid. This data shows that cloned oriT fragment is functional in the pPL1 plasmid. Plasmid pPL1 can be used for DNA cloning, gene inactivation by homologous recombination. Cloned genes or part of the gene cloned

could be then transferred by conjugation in to another host. Another advantage of this vector is conjugative transposon that can be excised from vector and inserted randomly in vivo in to another DNA molecule by purified Tn 5 transposase.

Construction of a conjugative, site specific integrative plasmid- transposon for horizontal

5   gene transfer between *E. coli* and *Streptomyces* strains.

PPL1 plasmid was used to clone an integrase gene from phage φC31, resulting a plasmid pPAOI6 (Figure 3a, 3b). We tested several clones for horizontal transfer between *E. coli* S17.1 strain and *Streptomyces lividans* TK24 strain. The best transfer was obtained for plasmid pPAOI6 were orientation of integrase gene is in opposite orientation to the gene

10   for resistance to apramycine. Conjugative transfer of pPAOI6 gene in to *S. lividans* strain is confirmed by resistance to apramycine or G418. Additional confirmation of transfer was obtained using PCR method. We were able to amplify 2 kb insert using specific primers for the φC31 integrase gene, and no PCR amplification was obtained for control *S. lividans* TK24 strain.

15   pPAOI6 transposon was cloned into the *EcoRV* site of plasmid pGPS3 (New England Biolabs). This construction allows transposition not only by transposase Tn5 but also using Transposase ABC (New England Biolabs). Resulting plasmid pTn5-7AOI is shown on Figure 10.

The goal of these constructions was to produce transposons that is further used in

20   functional analysis of the metagenomic DNA library from the soil that were constructed in a laboratory (Figures 4a, 4b, 4c).

Functional analysis of the metagenomic DNA library from soil.

The fosmids library consists of 120 512 clones, containing ~ 40 kb inserts of soil DNA. The library contains approximately 4.8 Gbps of the DNA cloned from soil. Ten percents

25   of the library was screened by using a PCR approach for the presence of the genes involved in production of secondary metabolites (PKS). Using gene-module specific set of primers we were able to identified positive clones organized in microtiter plates (96 wells). Sequences (based on PCR products) obtained from fifteen randomly positive clones indicate that the DNA library contains very little sequence redundancy limited to

one and that the sequences were found to be new and very diverse in comparison to gene databases (data not schown).

Two fosmids DNA was prepared from two positive clones (FS3-124 and FS3-135) and analyzed by sequencing. DNA analysis *in silico* shows high G+C contents of 72% and

5   69% respectively of the cloned inserts and presence of cluster genes that could be involved in biosynthesis of secondary metabolites (Fig 5 a and b). No specific phenotype was observed for the two clones in *E coli*. We employed pPAOI6 transposon mutagenesis to produce conjugative mutants. Transposon mutants FS3-124::pPAOI6 and FS3-135::pPAOI6 were isolated using apramycine as selective antibiotic. Obtained

10 mutants were then tested on LB pates containing chloramphenicol. About 1% of the tested clones are chloramphenicol sensitive. These clones contain transposon inserted into locus encoding chloramphenicol resistance gene and not into cloned DNA insert. ApraR and ChloS transposon mutants are then used for horizontal gene transfer into *Streptomyces lividans* TK24 strain. Fosmid DNA was prepared from mutants and

15 transformed into *E. coli* S17.1 strain. Horizontal gene transfer between *E. coli* S17.1 and *Streptomyces lividans* was done due to inserted pPAOI6 transposon. Transconjugants of the *Streptomyces lividans* were tested by PCR to confirm gene transfer and integration of the conjugative fosmid in to *S. lividans* chromosome. Both transconjugants showed an increase in doubling time, morphological modifications and pigments production in

20 comparison to the control (Figure 6).

B - From *E.coli* to *Bacillus subtilis*

Plasmid pPSB was constructed as follows : A part of amyE gene from *B subtilis* was

25 amplified by PCR using primers amyE-BamHI : atcgcaggatcctgaggactctcgaacccg (SEQ ID NO: 15) and amyE-EcoRI : cgactgaattcagatctagcgtgtaaattccgtctgc (SEQ ID NO: 16). DNA fragment was digested by *Eco*RI and *Bam*HI restriction enzymes and ligated into *Eco*RI, *Bgl*II site of pPAOI6 plasmid. Transposable element contains Tn<amyE-int φC31-oriT-apra> is shown on figure 7 with plasmid pPSB.

30

33

Plasmid pPSBery (Figure 8) was constructed by cloning erm AM gene into pPSB plasmid. A 1140 bp *Sau*3A DNA fragment containing ery AM gene with his own promoter was cloned from plasmid pMUTIN (Wagner *et al.* 1998) into the Bam HI site of plasmid pPSB. Orientation of ery gene was confirmed by sequencing. Transposable

5   element contains Tn < amyE-int ϕC31-ery-oriT-apra> (Figure 8).

Plasmid pPSBery-ΔI was obtained after Sma I digestion and self-ligation of core plasmid (Figure 9). In this construction, ϕC31 integrase gene was deleted from transposable element. New transposon is Tn<amyE-ery-oriT-apra> (Figure 9). All transposons could

10   be released as linear DNA by PvuII digestion from plasmids mentioned above and transposed by transposase Tn 5 (Epicentre) in vitro.

The selection of transposed elements was done using 100 μg/ml erythromycine or 40μg/ml apramycine in *E coli* or 0.3 μg/ml erythromycine in *B subtilis*. DNA was

15   transformed by electro-transformation into electrocompetent *E coli* strains or by competence into *B. subtilis*. Integration of imported DNA into amy E locus of *B. subtilis* chromosome was confirmed using pPSBery-ΔI plasmid. Integration was confirmed by PCR using plasmid-specific and amyE locus-specific primers. Fifteen eryR *B. subtilis* clones were tested by PCR. All transformants showed integration at amyE locus of *B*

20   *subtilis* chromosome, confirming the functionality of the method and constructs.

34

References

Brau, B., Pliz, U., and Pipersberg, W. 1984. Genes for gentamicin-(3)-N-acetyltransferases III and IV: I. Nucleotide sequence of the AAC(3)-IV gene and possible
5   involvement of an IS 140 element in its expression. Mol Gen Genet. 193 (1), 179-187.

Guiney, D.G., Yakobson, E., (1983). Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. Proc Natl Acad Sci USA, 80: 3595-8.

10   T. Kaeberlein, K. Lewis, Epstein S. S. 2002. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. Science 296, 1127-1129.

Mazodier, P., Thompson, C., 1989. Intergeneric conjugation between E. coli and Streptomyces species. J. Bacteriol. 171, 3583-3585.
15   Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science 276: 734-740.

Pridham, T.G., Anderson P., Foley, C., Lindenfelser, L.A., Hesseltine, C.W., and Benetdict R.C. (1957). A selection of media for maintenance and taxonomic study of
20   *streptomyces*. Antibiotics Annual 1956-1957, 947-953.

Voeykova T, Emelyanova L, Tabakov V, Mkrtumyan N. 1998. Transfer of plasmid pTO1 from *Escherichia coli* to various representatives of the order *Actinomycetales* by intergeneric conjugation. FEMS Microbiol Lett. 162(1):47-52.
25

Sambrook, J., Fritsch E.F. and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.

Torsvik, V., Goksoyr, J. Daae, F. L. 1990. High diversity in DNA of soil bacteria. Appl.
30   Environ Microbiol. 56, 782-787.

35

Torsvik, V., Sorheim, R. and Goksoyr, J. Total bacterial diversity in soil and sediment communities a review. J. Industr. Microbiol. 17, 170-178.

Torsvik, V., Daae, F.L, Sandaa, R.A., Overas, L. 1998. Novel techniques for analysing
5   microbial diversity in natural and perturbed environments. J. Biotechnol. 64. 53-62.

Wagner, V., Dervyn, E. and Ehrlich S. D. 1998. A vector for systematic gene inactivationin *Bacillus subtilis*. Microbiology. 144, 3097 – 3104.

10   Zaehner, H. and Fiedler, F. P. 1995. Fifty years of antimicrobials: past perspectives and future trends. In The need for New antibiotics: Possible Ways Forward. Fifty –Third Symposium of the Society for General Microbiology, Bath. England, UK, (Hunter, P.A., Darby, G.K. & Russell, N.J., eds), pp. 67-84.

15   Zechner, E.L., de la Cruz, F., Eisenbrandt, R., Grahn, A.M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C.M., Wilkins, B.M. and Zatyka, M. (2000) Conjugative-DNA transfer process. In Thomas, C.M. (ed.) The horizontal gene pool. Harwood Academic Publishers.

<u>CLAIMS</u>

1. A method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer and integration of said vectors and/or of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.

2. The method of claim 1, wherein the library comprises a plurality of unknown polynucleotides.

3. The method of claim 1 or 2, wherein the library comprises a plurality of environmental DNA fragments.

4. The method of any one of the preceding claims, wherein the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids, fosmids, P1 or BAC.

5. The method of any one of the preceding claims, wherein the selected vectors are modified by targeted insertion, into the vector, of a target polynucleotide construct.

6. The method of claim 5, wherein the targeted insertion is performed in a region of the vector distinct from the polynucleotide.

7. The method of claim 5 or 6, wherein the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.

8. The method of claim 7, wherein the origin of transfer is functional in *E. coli* host cells.

●

9. The method of claim 8, wherein the origin of transfer is selected from RP4, pTiC58, F, RSF1010, ColE1 and R6K(α).

10. The method of any one of claims 5 to 9, wherein the target polynucleotide construct
5   comprises an integrase functional in the selected host cell.

11. The method of claim 10, wherein the integrase is selected from φC31.

12. The method of any one of claims 6 to 11, wherein the target polynucleotide construct
10   comprises a transcriptional promoter functional in the selected host cell.

13. The method of any one of claims 5 to 12, wherein the target polynucleotide construct comprises a transposable nucleic acid construct.

15   14. The method of claim 13, wherein the transposable nucleic acid comprises, flanked by two inverted repeats, the target polynucleotide construct and a marker gene.

15. The method of any one of the preceding claims, wherein the cloning vector comprises a first marker gene and wherein, in step ii), the cloning vector is modified by:
20   . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and
. selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.
25

16. The method of any one of the preceding claims, wherein, in step (i), the cloning vectors which contain a polynucleotide having a particular characteristic are selected by molecular screening.

30   17. The method of any one of the preceding claims, wherein, in step (iii), the modified cloning vectors are transferred into the selected host cell by conjugative transfer.

18. The method of any one of the preceding claims, wherein, in step (iii), polynucleotides are analysed by determining the phenotype or properties of the host cell upon transfer or expression of the modified vector.

5

19. A method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E.coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

15

20. A transposable nucleic acid construct, wherein said construct comprises an origin of transfer and elements for integration and selection in a selected host cell genome flanked by two inverted repeats.

20   21. A library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing transfer and integration of the environmental DNA into a selected host cell distinct from

25   *E. coli.*

22. A polynucleotide sequence comprising all or part of SEQ ID NOs: 1 or 2, or of their complementary strand.

30   23. An oligonucleotide comprising SEQ ID NO: 3 or 4.
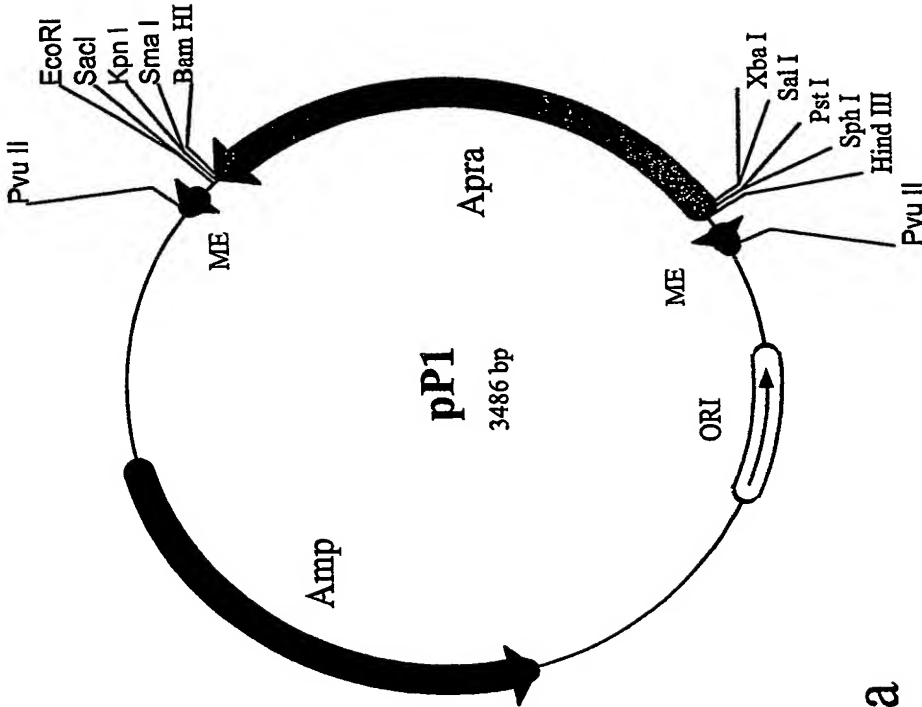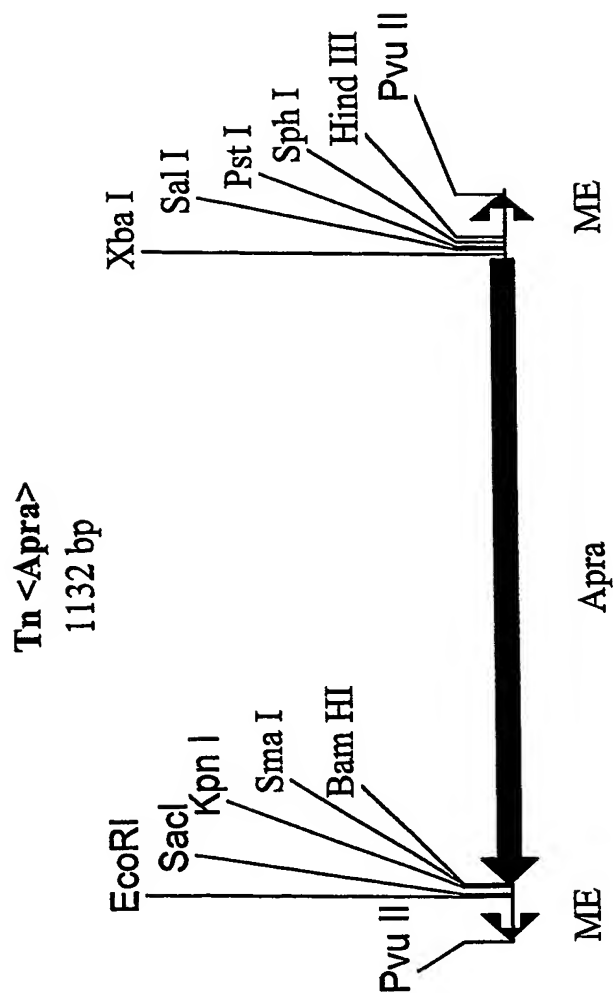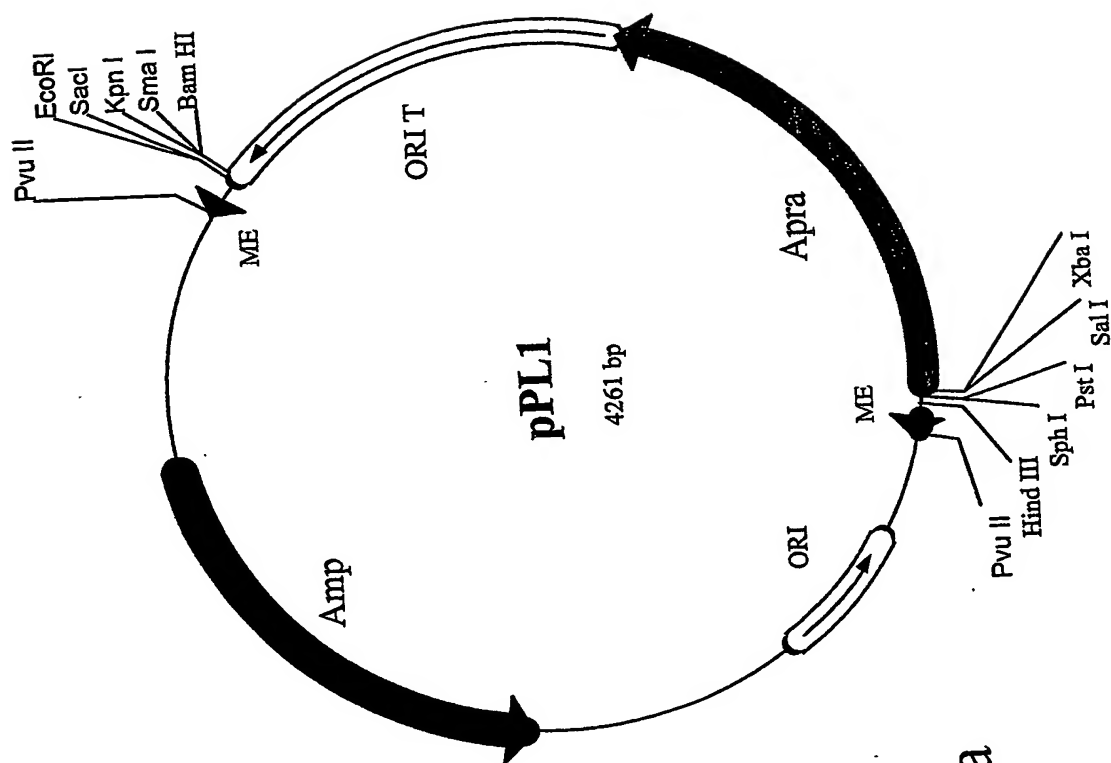
1/39

Figure 1a

Figure 1b

Figure 2a

4/39

Tn <Apra-oriT>
1907 bp

XbaI
SalI
PstI
SphI
Hind III
Pvu II

ME

Apra

ORI T

ME

EcoRI
SacI
KpnI
SmaI
Bam HI
Pvu II

Figure 2b

Figure 3a

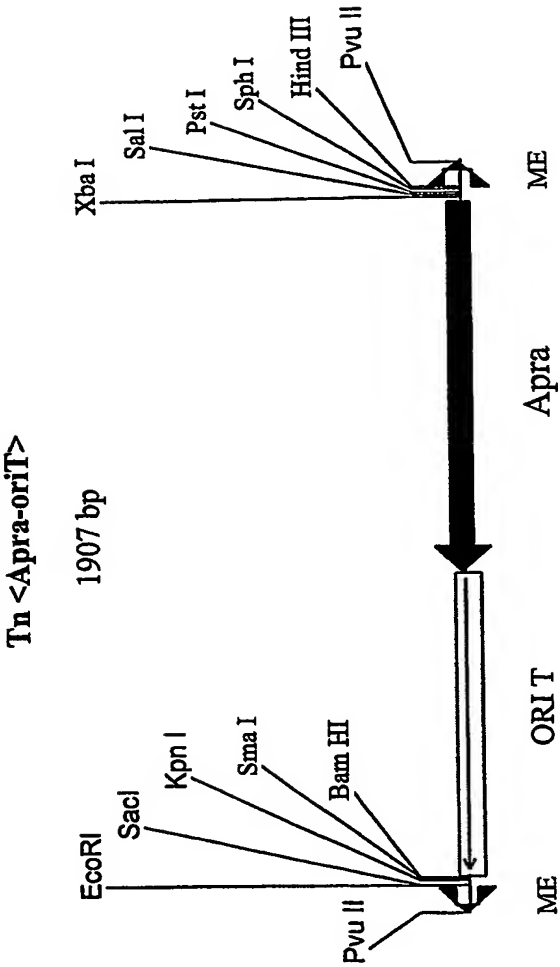Figure 3b

7/39

cloned insert

A gene

cloning vectors

Plasmids
BAC
Cosmide

X gene for selection
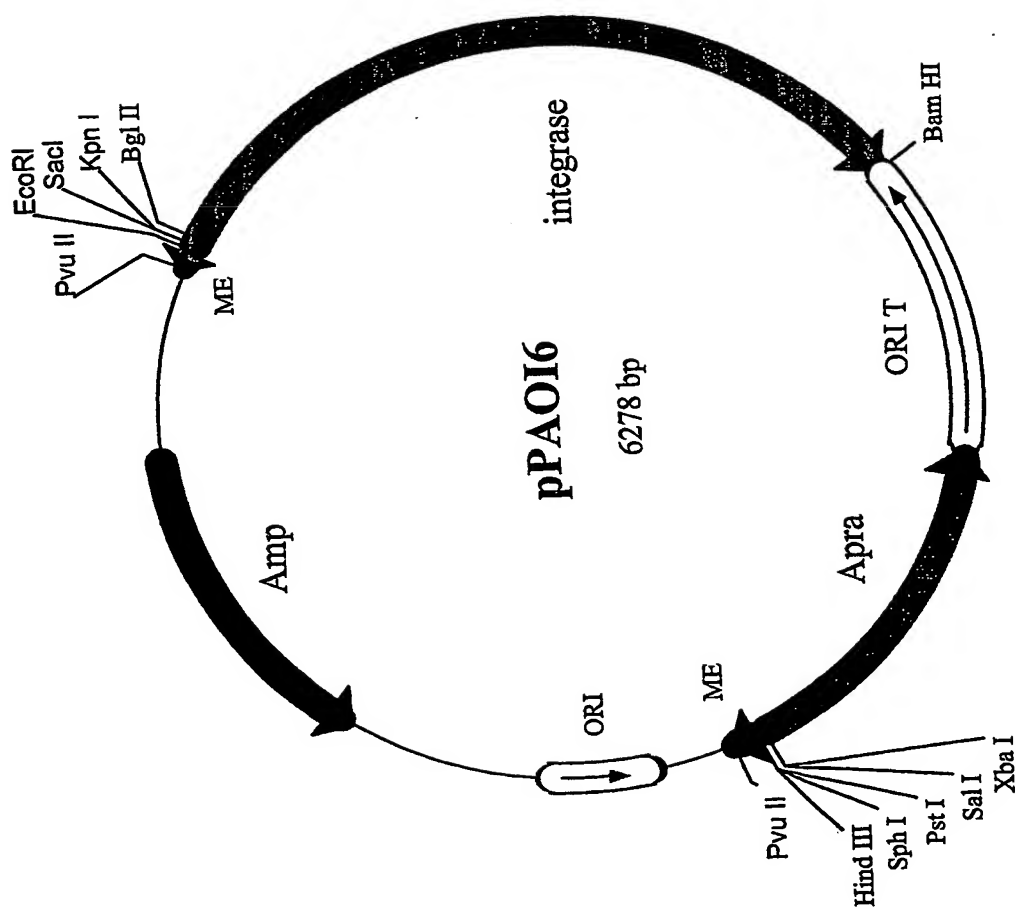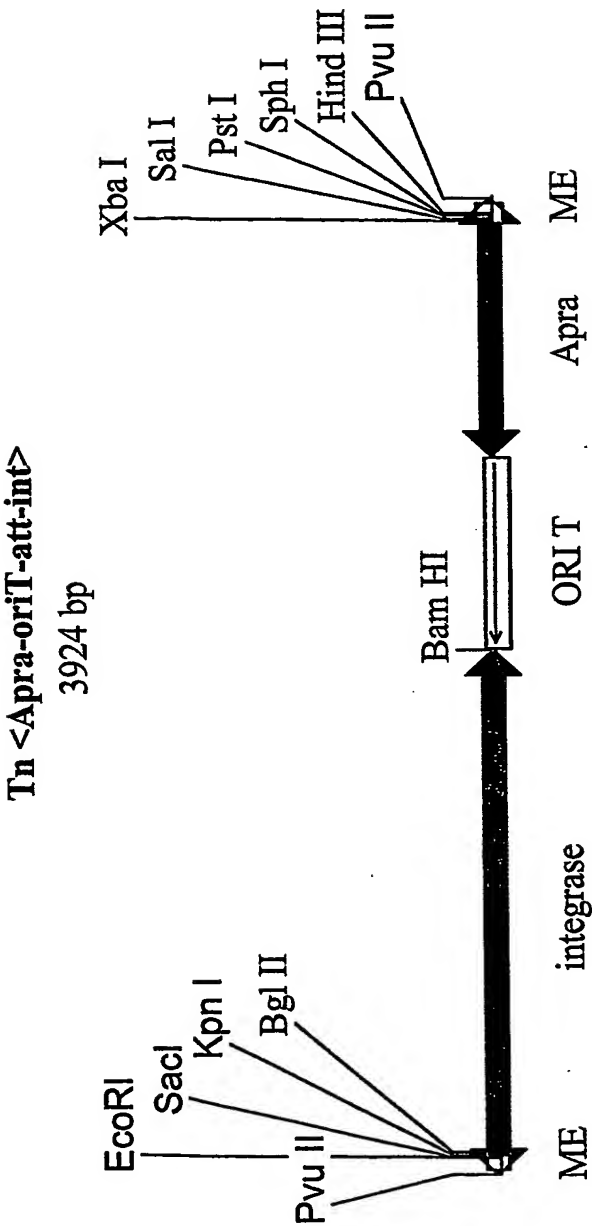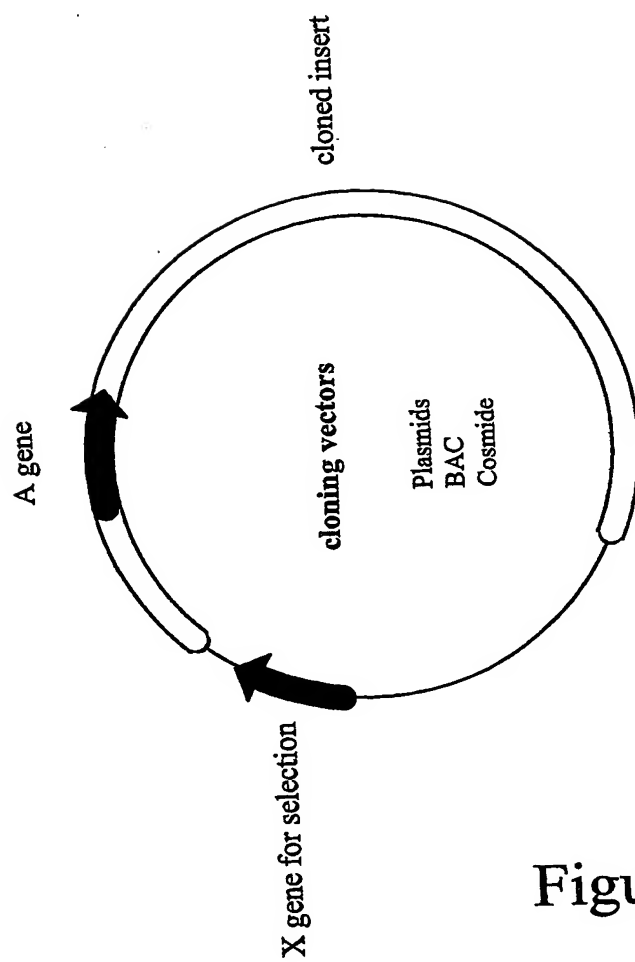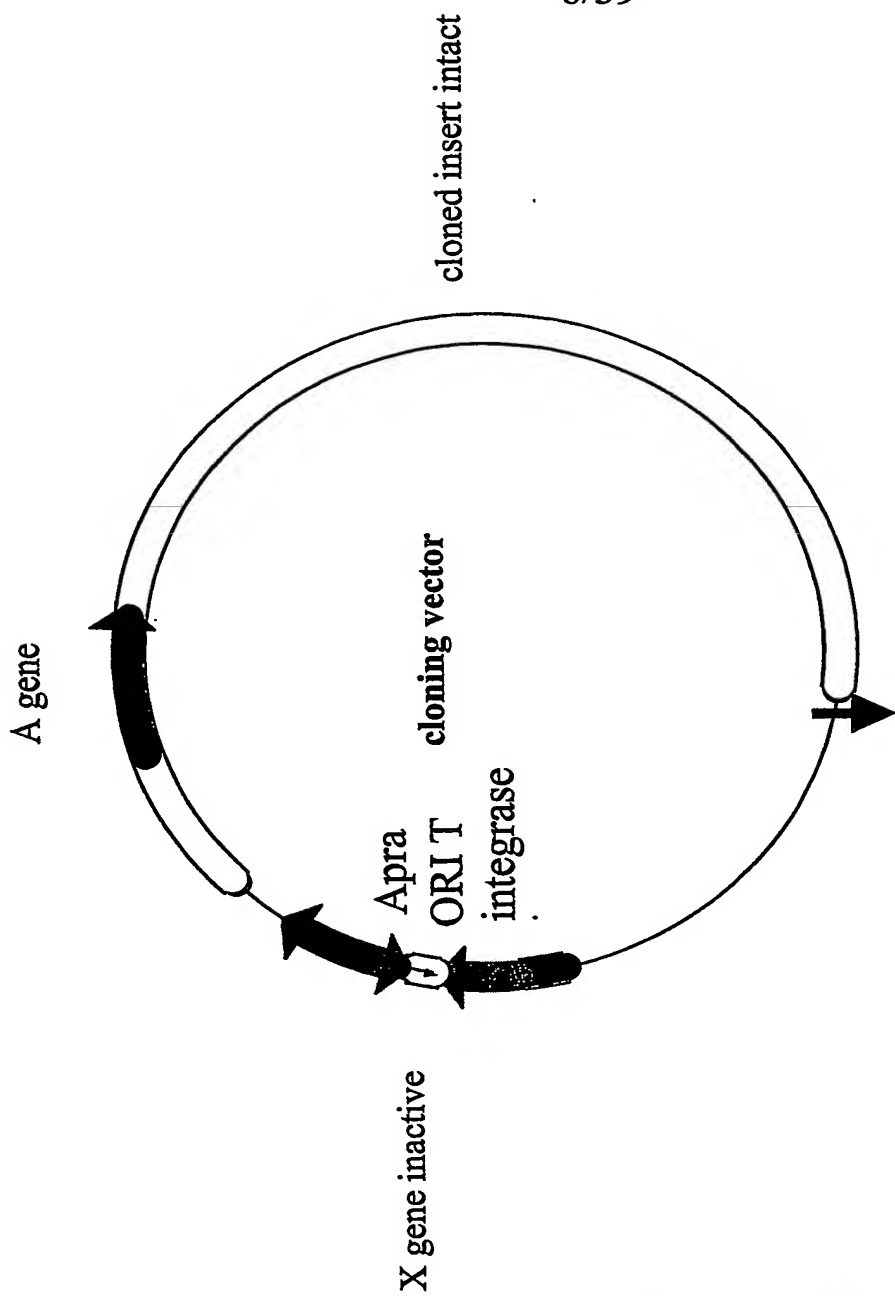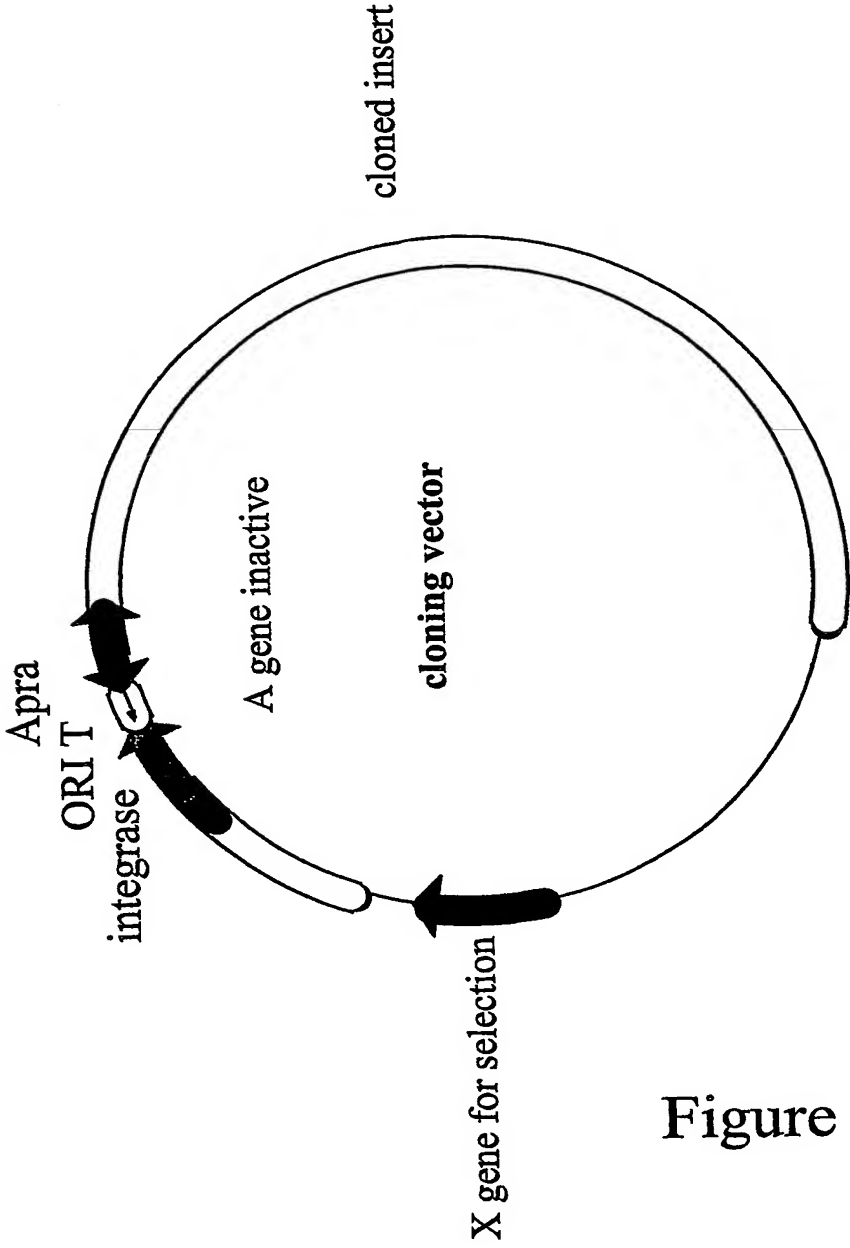
Figure 4a

Figure 4b

Figure 4c

Figure 5a

CDS	76..1134 /note="ABC_transportr" /gene="TAP2 PROTEIN" /blastp_match="Oryzias latipes" /blast_score= 0.002
CDS	complement(1096..2430) /note="none" /blastp_match="Anabaena sp" /gene="ALR1117 PROTEIN" /blast_score=2e-18
CDS	1178..1624 /note="Gram_pos_anchor" /gene="CELL WALL SURFACE ANCHOR" /blast_score=1e-04 /blastp_match="Streptococcus pneumoniae"
CDS	complement(2506..3567) /note="CONSERVED" /gene="HYPOTHETICAL PROTEIN" /blast_score=0.019 /blastp_match="Deinococcus radiodurans"
CDS	complement(2906..4222) /note="glycosyl transferase" /gene="lipopolysaccharide" /blast_score=2e-23
CDS	complement(4092..5321) /note="glycosyl transferase" /gene="glycosyl transferase" /blast_score=1e-15
CDS	complement(6337..8502) /note="PUTATIVE" /gene="GLUTAMINE AMIDOTRANSFERASE" /blastp_match=" Bordetella bronchiseptica" /blast_score=1e-16
CDS	complement(8181..9530) /note="none" /gene="MEMBRANE PROTEIN" /blast_score=0.035
CDS	complement(9531..10721) /note="NOEC Transmembrane" /gene="NODULATION PROTEIN" /blastp_match="Azorhizobium caulinodans" /blast_score=3e-07
CDS	complement(10504..11274) /note="PUTATIVE" /gene="HYDROLASE" /blastp_match="Streptomyces coelicolor" /blast_score=4e-14
CDS	12874..13689 /note="HYPOTHETICAL Meth-transf" /gene="PROTEIN PA1088" /blast_score=2e-06 /blastp_match=" Pseudomonas aeruginosa"
CDS	14195..15976 /note="PUTATIVE, Glyco_transf" /note="" /gene="LIPOPOLYSACCHARIDE BIOSYNTHESIS" /blast_score=2e-06 /blastp_match="Vibrio cholerae"
CDS	15427..16512 /note="PATHWAY: INNER CORE LIPOPOLYSACCHARIDE BIOSYNTHESIS" /gene="PHOSPHOHEPTOSE ISOMERASE " /blast_score=3e-17 /blastp_match="Helicobacter pylori"
CDS	15579..16253

```
                              /note="none"
                              /gene="PHOSPHOHEPTOSE ISOMERASE"
                              /blast_score=2e-22
                              /blastp_match="Neisseria meningitidis"
        CDS                   complement(16505..17656)
                              /note="BIOSYNTHESIS PUTATIVE"
                              /gene="LIPOPOLYSACCHARIDE"
                              /blast_score=2e-17
                              /blastp_match="Thermotoga maritima"
                              /pfam_match="Glycos_transf_1"
        CDS                   complement(17657..18697)
                              /note="none"
                              /gene="ALR3073 PROTEIN"
                              /blast_score=6e-27
                              /blastp_match="Anabaena sp"
        CDS                   complement(18615..19304)
                              /note="none"
                              /gene="ALR4487 PROTEIN"
                              /blast_score=8e-07
                              /blastp_match="Anabaena sp"
        CDS                   complement(19301..20596)
                              /note="ATP_GTP_A"
                              /gene="ABC TRANSPORTER"
                              /blast_score=3e-61
                              /blastp_match="Synechocystis sp"
        CDS                   complement(20535..21476)
                              /note="PERMEASE COMPONENT"
                              /gene="POLYSACCHARIDE ABC TRANSPORTER"
                              /blast_score=6e-41
                              /blastp_match="Clostridium acetobutylicum"
        CDS                   complement(22025..22951)
                              /note="involved in the synthesis of a polysaccharide
                              capsule ?"
                              /gene="32.3 KDA PROTEIN"
                              /blast_score=3e-17
                              /blastp_match="Sphingomonas sp"
        CDS                   23155..26523
                              /note="peptide syntase"
                              /gene="mcyA, mcyB and mcyC"
                              /blastp_match="Microcystis aeruginosa"
                              /blast_score=0.0
        CDS                   26409..34433
                              /note="polyketide syntase et peptide syntase"
                              /gene="mcyD, mcyE, mcyF and mcyG"
                              /blastp_match="Microcystis aeruginosa"
        CDS                   34418..37500
                              /note="CYSTATIN"
                              /gene="PEPTIDE SYNTHETASE"
                              /blast_score=0.0
                              /blastp_match="Anabaena sp"
        CDS                   35359..37500
                              /note="gene cluster"
                              /gene="nostopeptolide biosynthetic"
                              /blast_score=2e-41
                              /blastp_match="Nostoc sp"
      Sequence 37500 BP; 6199 A; 12698 C; 12769 G; 5834 T; 0 other;

    gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt gctgcaaggc      60
    gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg acggccagtg     120
    aattgtaata cgactcacta tagggcgaat tcgagctcgg tacccgggga tcccacgtac     180
    cacgagctca tctggaagag cccgctccag acacccgagg aatggaagcg cggcgtcgtg     240
    agaccgtaca cgcagaagcg tctcgtcgcc ttcgttgact ggcttgcgac gacgatgaag     300
    ctggatctca ccaggatgtt cgccgcaggc agctcgatgg gcggatcggg cgcgatcatg     360
    ctcgcgattc gctatcccgc acgattcgcg tggaccgtgt cgtgggtcgg cgtccacgtg     420
    cccgccgact ctccgcaatt cacatcgtcg tacgagctgg tgtacggccg gcccgacttg     480
    aaggtgccgt tcgagaacgg cacgccggtc tgggatcatt ttagtgacgt ctggtacctg     540
    cggcagcatc cggagcagga catcgggttc atcacgttct cgaacggcaa gaacgactcg     600
    gcgatcggct ggcgccaggc cgtcgaattc ctgaagacgc tgcaggagac caggcaaccg     660
    cacctgttcg tctggggcca ggagggacac ggccagcgcg cgaagatgcc ggaaggtggc     720
```

```
ggcgaacggg agatgccgct cgacctcaga acgaaccaga gcctgccggc gttcagccgt    780
tgcacgctcg acgacgatcc gggagacggg tcggattcga gcggcgtccc ggccgggcag    840
atcaactcct acgtcacctg gcagcctgac tcgatcgtcg acgaacccgg ccgatggagc    900
gtcgtcatca agctggcgaa tcgaccgccg cggtcgcccg ccgcagtcga cgtcacgccg    960
cggcggctgc agcagttcaa gctcaagccc ggcgatcagg tgacatggac gaactccgcg   1020
ggcaacaccg tcgtgcagcg cggcgaagcg gtcgcggacc ggtgggggtct cgtcacgctc   1080
ccgcagatgc aggtctcaac gggtgggaaac aggatcctgg tgagcaggaa ctaggcgggc   1140
ggcgtcctcg acgagccggc gcatcggcgc cagttgatgt gcccacgaaa gctggctgag   1200
gtccgcgcgc gcggcgatgt cccgtcgcgc ctgcggcggc tttgccagca aatccagcac   1260
agccgccgg aacgcctcga ctgattccgc cacggagcac gccggccgtg cgatggtggg   1320
aagtccccct gcgacggccg gggtcacgac gttcgggagc cccgcggcaa tcgcttcgag   1380
cactttgttc tgcacgcctc tcgcgatcgc aatcggcgcg accgcgatcg ccgatcgcca   1440
gagatactcg cgtacctccg ggaccgcgcc ggtcacttca atcgtgggat cggccgcgag   1500
tgcccgcacc gcggcggtcg ggtccatgcc cacgagcatc agccgcgcgt ccgggcgggc   1560
gcgaaccacg gacggccaga cctctcgcgc gaaccagagc gccgcgcgct cgttgggctc   1620
gtagccgaac acgccgcaga agacgacccg cgcttcggcg gccggtggtc ctttcggcgc   1680
aaacatctcg acgtcgatgc cgttcggcac gaccgacgcc gtgaggccgc ccgtgacttc   1740
ttcgagcagc aggcgctcgc gctcgctcac gaccgtggtc gccgccgcgc gccgcatcgc   1800
gactttctcg aagcggagca agcggccggc ctcgcggcgg aacagccagc ctttcggcca   1860
cgccgccgtc gacgccagcg cgcgccattt ctcggagtcg acgtccacca tgtccaggac   1920
gaaggggatc ccgcccagga tcggctcgca ggcgtagcgg gccatcccgc tgcagtacgc   1980
gagcaccgcg tcgggccgcg tcgcggcaat ccgtcgttcg agcacctggt agatctccgg   2040
cgaatgcagc aggacgtgcg tgaggggtcg atcgcccggc agcgccagac cgctcgcgat   2100
gagattccgg ccgcggcgga cgcgcacgac gtcggtcgac gccgtcacgc cggcgagatc   2160
cgctcgatgc gaccactcct cgtcgtcgtg cgcgagcgag acgaggtgca cgtcggcgga   2220
tcgtgcgagc gtgtgaatga gatgaaacgc gcgaatgcga tcgccccggt tcggggcgta   2280
gggcagccgg tgtgtgagca gaaggactct caccgtgcgg catccgatcg ctctcggatg   2340
acgcgggccg gtacgcccgc ggcgatcacg cgggcgggaa tcggcgcggt cacgaccgac   2400
cccgccgcaa tcacggaatc cggaccgacg tcagccatga ccacggcggc gctgccgatc   2460
caggtgccgg cgccgatccg caccatccgc ggtgtgcctg gctgctcacg gatcggccgc   2520
gtgaggtcgc tcgtgccgtg cgtgtccggt ccgctgggga tgtggacgcc ggcgccgacc   2580
aggacatcgc gctccaggtg gacccagccg agatggcatc cgggtcccac gtacacatgc   2640
tcgtcgagct tcgcgccggt ctgcgaaaag atggtgccga aggcgaccgt gaccgacggg   2700
tcgcagtgtg caagcacgcg gccgaggaaa gccgcgcgaa ggtattgtcc gatgacgccg   2760
gggatgagcg agagccactg cgtcgaccct tcgagggccc ggtcggcgcc gaggaagggc   2820
cgccgcacgg agtaggagag caaggcgggc aggacgagca ccagcgcgat gccacgggcc   2880
agtcccttgc cggcgtcctt gaggctcacg cccactgcag cgcgcgcccc gcgtcggacc   2940
gggctccgta tcgtgtgacg agcgagtcgt agagcgcttc cactttgcgc atccgccgct   3000
cgaacgataa ctcttcttcg atccggcgtc gcgccgcaat cgcacgggtg cgtccggccg   3060
cgggattggc catcgcctgc tcgatcgcgt ccgcgaggcc tgacgcccgg ccgggctgaa   3120
tgaccagggc gtgcgtgccg ttcgacacga gctcggccgt gcccccggcc gcggtcgcga   3180
cgatgggtgt ctcgaacgcc atggcttcga gcacggcgtt gggcgtccct cgtagtccg    3240
.acgactgcac gagcaggtcc agcccgtgat ggacggccga gacgtcggac gtgtggccga   3300
ggaaacgcca ggttccgggc ccgagctcgc gggcggcctg ttcctcgagc ggacgccgga   3360
ggctgccgtc gccggcgatg acgagacgga ggcgcggcca cttcgcctga agggcggcac   3420
aggcttcgat gagcaagtcg aatcgtttct gcggctcgag acggccaacc gcgccgatga   3480
cgaagtcgct ggcctcgaat cccagtcctt cacgcgccgc ggcctcagat cgtggatcgc   3540
gtcggaaagc atggtggtcg atgccattcg ggatcgtgac gacccgatcc ggccgtccgc   3600
ctttggccag gagctcccgc cgaatctgtt cggacacggc aatgacggca gggaagcggg   3660
cgagaatccg gcgatcgagc gggtagtaga cggtgcgttc gcgaaacgaa tggccggtcc   3720
atccatgggc cgtggccagc ggaatcacgc gctcggcgcg tcccaggagc agggcgagga   3780
gatcggtctt gtactcgtgc gcgtgcacga tgtcgatgcg cagatctttc acgagcgcgc   3840
ggagcttcga ccagattccg gtgtcgaacg aatgcttctc ccggacctcg acgtagtcga   3900
tgccggccgc tttcgcccgt tcgtccatgc cgaacaccgg atcgcgctgg tcgcggatgt   3960
agcagacggt gatcgcgtac ttggagcgat cggcctgcgc cgttccgagc agaatcgtct   4020
tctccggtcc gccgcccgtg ccccgcacgc tgcggagctc gagcacgcgc acgggacgcg   4080
cgattcccgg cctaggcgcg gcgtctggca gcatggcccc cccggccgac gagcgcgatg   4140
gcccgggcga tatcgtacgt gccggccacg acgagcgact gtgcgatctc gacgagtccc   4200
actgcgccgc gccggcgccg gatcatcgcg gctgcgcgca agcggcgaa accggcaagc    4260
agcaggagcg cggtaccgac aatccaccac ccgccgcgct gccacagcgc gaggcccagc   4320
ggaatcgaga cgagcgcgac cagatcgagg accgggatga tcacgctggg cagatcccgc   4380
atggcgagcc ggtgacggaa gctcacgcgc aggttgtcgc ggccgcgcca cagctcgccg   4440
aaaaacacgg cgcgaagcga tcggggatcg ccctggtgca cgctgcggag ccgcgggtcc   4500
gacaggatgc ggccgccgcc ccggcgcagc cgctggcaga tatccacgtc ttcacaggtc   4560
tcgaggctgg tatcgaagcc accggcgcgg tcgaaggcct cgcgcttcac ggcaagattg   4620
ccgctcggga gccagtccac gtcttcggtg ctctgcccac gacggcgaaa cgagtcgtac   4680
gcgcgctgca cccagttgcc gtcaccggga gcttcgtact gcgcgccggc agcgacgacc   4740
cccgggggtct ggagcagatc gacggcgacc accggccagc gcgggtcgat ctcatgatcg   4800
gcgtcaacga acgccagcac gtcgccctcc gccgaccggg ccgcgcggtt gcgaagcgtc   4860
```

```
gccacgttca cgtcgggcag cgacagcact cgcgcggatc cgcgctctgc cacggcacgc    4920
gagtcgtcgg tcgacccgtt gtccgccacg acgatctcgt accagccagc cggcgcctcg    4980
gtgcgatgaa tcgacgcgag gcaacggctg aggtgcgccg cgccgttctt gacgggatg     5040
acgaacgaca cgcgcggcgc cacggcaatt ttcgcggtca tcgatttcgt agcgcggccg     5100
gctttcgtag cgcggcccgg gttgggggcc gcgagaactc gaacgcgccc agatccggcg    5160
cccctgctct cggccgccga tcgaaatcga gcggcgcctg cggaatctcg atgccggcgt    5220
cgatcgcctc cctggcggct gccgacaggt gcagatcgca cgaggccgcg tttgcgaacc    5280
actctgcggc ggcgttcgtg acgttgttcg tcgcctccgc ctcaccgccg ttacggcgca     5340
ggatctgccg gctcgtcagg ttgttgcgga tgagccgtt cgtcacggga aaccggacgc      5400
cgatcgtcca cggcgtcgcc gtgccgagcg tgaagacgct gttgtgctcg atccggaagt     5460
cggtggacgc gttggcttcg atcgcttcgt cggcccattc gttcaggttg cagatcacgt     5520
tgttccgcac gatgccgcgc tggtgatcga tctgcttctc gccgtcgcgt gcgtaggccg     5580
tgatccccgc ggccaggccg atggcgatgc ccctgaacga atcgacgatg acgttccgtt     5640
cgatgagcgt gtcctgcgag ttcgcccagg cgaggatcgc cgggccggat cgccagccgc     5700
ccgattgcgg gccgcgaatc cgcaggaaga cgttgtcgcg gatgacccac ccccgggccg     5760
ccaggatgtc cacgccgtcg gtgtaatcgg acggcgcgct cgtcgtgtac tcgaagcgcg     5820
agcacgccac gaggccatcg tccgcgaagc gtccgtcagt gccgacgctg cctttgagca     5880
attgctgtcc cgcatcgatg agctggacgt tgtgcaccgt ggctcgggac gcgccaagct     5940
cgccgcgcac ctgaatcgcg tggtacccga cgtggccgac cgtcagatcc gcaatcgtca     6000
cgtcagtggc gccaaccgag agcgccacgc cgatcgcgtc gccggtcatg ccgccgccgc     6060
gaatgacgac gcggtcccga tcgccggagc ggctccggag cacggtgccg gggacgttga     6120
tgctggccat ccggtcgagc cggtactcgc cgttctcgag cagaatcgtg gtgcgcggct     6180
tgacgcgttc gagcgcctgc agcagctcgg acgtccgccg cacggtgacg acctgtttcg     6240
tcgccggcgc cgcccatccg cagaagtgcg gcaccggcgc ggcggcctgg ccggacggca     6300
gaagcgccac gagtgcgccg gccacgatga ggcgagtcat gccgccaccg tgtcgacgtc     6360
gtgttcgcgc aaccagagct cgaggatgag caggatttga aggaggtagc cgcgcgactg     6420
tctcccgtga cgcgtgtcgt cgatgatggc gcgcaacgtc tcggggcgca ggaacatccg     6480
cgtgcgcgcc tccggggcga acagctcgcg ctcgacggct tcgagcagtc gcgtctgcag     6540
ccagcgatcg aagtcgtgat agtgacggta cccaggcgcg ttcaggcgct tcagcagggt     6600
gttgaccttg tcgaacacct tctccgccag cggaccggcg tcgccgcgcg cgccggtgtt     6660
cgagttccgc accttgagca gcgccgggtt gccggcctgc gtgatcgcgc gatgaagcgc     6720
cgtgtcgtcg cgccatgccg gatcgccgct cagcaggacc tgcaggaact cgcgatccat     6780
gaacggcagc cgcacgtcca cgacggtgcg gaagagatcg agcgagggaa tcgtgaagcg     6840
ccgatgatgc tcacgaagat agagatagct gctgagctcc gcgggtgaga gctccacggt     6900
gtcgaggagc gcgtgcagcg agtcggccgc gccccgtccg gagaggcgcg cagcgtctgc     6960
tgtgaagagc gtgtccagcg gcaggccgct gctgatgtag ttggcgcgcg acgtcatgta     7020
gtcgatgaac tcgccgcggc ccgtcatcgc gaacacgtgg ccgtcggtgt ggagcggcca     7080
cgccaggctc gtcttggcga gctctcccgc gtggccgcgc agcaggacct cgataccgga     7140
gtcccgcagg aaatcgatcg ccagcatctc ggtcagtccg tggctcagat acatgccatc     7200
ggtgagcgag accatgcgcg cctggttcgg cagaaagtcc ttgaggtacc ggtcgtcgag     7260
ctcgaagaag cggtggtcgg taccggtgag cttcgacagc tgttgcgcga tcacctgatc     7320
ggcgcagccg cgacgcccga gcgtgtacgt gatgagcccg ccaacaccgt tcaccgcact     7380
cagcacggcg cggctgtcga ggccgccgga gagcgacagc ccgaagcgat gcctgcccga     7440
gagcgacttc tcgaccgcgc gcctgaacgt cgtgtgcacc gcgtggacgt acgcgtcctt     7500
cgtcgtacgc ggtccgcggt agagatcggc cggctcgaaa taccgggaca cgctcgcggc     7560
gcctgacggc cagtcgtacg cgagcacggc gccggggtcg aggagcgtga cgcccgccgc     7620
caggtcttg tcgcccagga cgaagcccag cgtgacgtaa tcggcggccg cccggaggct     7680
gagagccggt ttgcgcgccg aggtcctgag cacggaagcc agatccgatc cgaacgacag     7740
tccggccgcg tcgacccgcc agtagatcgg atacgagccg aacggatccg ttgcaagcag     7800
cactcgctct cgcgccggat cgatcacggc gagcgaaaac gcgccgtcga gccgggcggc     7860
catggcaggc cccagccgtt cgtacaccaa agccaggaga tcggcgcgcg tcccgttcgc     7920
cgaaccaaac tcgctccgga gcgccgcttc gttgtagaga tcgccgcgaa agatcaccag     7980
cggccggctg acgtcgcgcc cggccgccac ctcgtcgagg aagccgagcg tgccgagccc     8040
gaccgcccat cggtccgacg gatctcgata gacgaccgtg cggcgcggcc tgcccgactt     8100
cgccgctgac gcggcgtcgg cgcggctgaa ggcggtagat gccggctcat cgcggcaaac     8160
gacgccgagg aggccgttca tcagacgacc ccgcgggcag acacgcgcga cggaacaaga     8220
ggtgtcgtct gagcggccgg aaccggccgc gcggcgtgcg ccgccaccat ccggttgagc     8280
gcaatcgaca agccgaggaa gtgccagagg atctcggtgt actgaaagat gacgaacgtt     8340
ccgcccacgc agaatgcaat gagggcggtc tggacggcca gcgcgtacgg cgccagctcg     8400
gccagggccg gatcccgctt cgagacccgg cggaccctgc ggcaggcgat gagcgccagg     8460
aggagagtga gcaggaacgc cgcgaggccc acgaacccaa gctcagccac gagaccgaac     8520
cacgcgctgt gaaccgacgg ccccttgccc caggcgccgt ccgacgtgtc gtagtcgttg     8580
taggcggccg tgaaggcatt gtgaccgacg ccgtaaatcg gccggtcatc ggccatgcgg     8640
agcgcgaccc tccagaagtg cagccggctc tgggccgatc gctgatcgcc aaccgccgtg     8700
tcgctcgcgt cgatgctcga gtcggacgag ctgatcgagc tcatgcgatc ccagaactct     8760
tccgtcatca cgagcgacat cgcgacgaag accgccgccc cggccacgag gacgagcatc     8820
ttgcgcttcg agtacaggaa atagacgagg ccaagcacgg cgagcgtgag gaagccgccc     8880
cgggagtacg tgctgatgcc gcgatagagc accccgatga gcagcacgcc cagcgcccac     8940
ttggcccatt tcctcgcttc cgtctgcatc agcgtgacga tgagcggcac gagcatgaac     9000
```

```
atgccgacgg cgacgccgtt gttgtcaccg agcgagtaga cgtcgttgaa gttcttgatg    9060
ccgggaatga acaggagctg accccagccc tgcttggcgg cctcgaaccc gagggagagc    9120
gagatcacga ggaagacgag ccgcaaccgc ttgacgtccg tcgtgaggac tgcgagcagg    9180
tacgtcatca cggtggactt catgaagtcg atcatgtagc cccacgcgaa gtcctggtag    9240
ggcgacgcca gcgtcgacga caggctgtgg agcccgaaga aggcgaggag cgccagccgc    9300
gcgtcgatcc gcagcgcctg cccggacagg aacgcgatcg cgagcgtgta catgcccgcc    9360
aggagcgaga tgttcatgtg gatgagaaag tcgctccaca cccagagctc cggccggaag    9420
tacgcgatga acaggtagaa cagaatcgcg tagaacggcc cgcgcagcgc gtggaaggcg    9480
ccgaaggcga ggagggcgag gacgaaagcg gctcgaagca ttgccgtgtg ttacctgtta    9540
tcggtcccga tggtccagag cggctgcgtc cgggtctcct ctacattgaa caccttgtag    9600
agggccggaa tcgaggtgaa catgagcacc acgaaaagaa tcgccgacac caccatgtag    9660
gcgaagaacc cccgctgctt gtagagcttc tccggattct gcacggggct gttcggcagc    9720
atcccgatgt gcaggtagta ggcgaacatg cccgccgcca ccggcgcgaa caggatcagc    9780
tcgaggtgat agcggacgat gaagatgccg ccgaagaggg gcgcagcaggt ggcgtagaac    9840
accatgctca cgagcagccg ctcctcgttg tagaaggcaa acgacttccg gtacgaggcc    9900
gcgaccgacg cgttgccgat gtgccggaac tccgcgtacc gcttcgtcgc catgaagaac    9960
gcgcccacca tccagtacga gatcacgagc gacatcggcg gcacgcgatc ctgaatcagc   10020
gggaaccagc cgaggaggag ccggacggcg ttgttcgccg actcgctcag cacgtcgagg   10080
tacggccact ccttcgtgcg aatgggcggc acgttgtacg tgacgccgag cacccagagc   10140
gccaggccg acagcgcgaa gtatctgttc acgagcagcg cgagcacgaa tccggccacg    10200
cccacgagga tccactcggt gtagcccgcc gccggcttga tcttcccgga gggcaccggc   10260
cggtgccgtt tctccggatg cagcagatct ctcgggccgt ccaggagctc gttcagcacg   10320
tagttgctcg aggcgatgag gcaggtcgcg gcgagcgcga gcgcgagcgg cgggacggcg   10380
gtccagccga agagctgcgg ttcgtagaag aacgccagca gcacgccgag cagcatgaac   10440
gcgttcttga accagtgatc gatccgcgcg atctgacgt acggccaaat ccgcgaaccg    10500
gagctagcg ctgacgacat cgttgagctc ctcgagccgc cggatcgcga accatccgcg    10560
gccggctgtg acgccgtggc ggccggcgat gagggcgcag gagagtcctg cggccatcgc    10620
tccttctccg tccacgtccg gtcgatcccc gacatacagg acctcccgag gtccgagtga    10680
ccacatctca cacgccacat ggaagccgcg cgggtgcggc ttcagggcgt tcacgttgct    10740
ggcggtcgtg cagagcgcca gcgaaaaatg ctgcgacgtg ccgagcgcgg cgagcttgct    10800
ctcgggcgcg tagtccgaca gcacgccgag cctgagcccc gccgcgcgaa gcgagtcgag   10860
cgcgccgagc aaccccgggc gccggcacca ccgcagatac ttgagcggcc ggcgcaccat   10920
ccattcgttc accgtcgccg ccacgacctc gttgtccagc ccgagcgcct gcgcggtgcg    10980
cgcgatctgg cgcgccgcga gcggctcgtc ggccgcgccc aggcgccgga gatcctcatg    11040
cgcccggcgg tactcccgca cgatgcgcgc cgtgtccact ccgcgccgcc accggaacgt   11100
cacgagcggc gtggcggcga gctccgctgc catggcggcg cgcaacggac cctgccggta   11160
cagcgtgccg tccacgtcga acagcacggc cttgaccccg cgcaacggct tcactctgcg    11220
aggcccagca tccgcgacac cc tttgacgcgt cgcgaaggaa cacgtcacga           11280
acccacacgg cagccagccc cgccaacagc accgccgctg ccgcgcccgc tgccgtccgc    11340
cagctcccgc gccaggcgcg cgtgtcgatc gcggcggcgc cgtagcagag cagaatcggg    11400
acgagcggca ggtgataccg gggatgtccg aagacgatcg cgtgaatgcc gctcatgaac    11460
acgatcaggc tcaggagcag caggtgcgcg cgccagttga ccggccgcgc gaacgagatg    11520
ccgatgacgc cgagcaccat gaccgccgcg tacgatccca tgatggacac cgccccggcg    11580
aagaacagcc accgcggcgg acggtagaac ccatttccga cgccggcgat gaagtcgcgc   11640
tccagccccc agaagtctgc gaacttcagc accgaacggc gaagcgtcgt tcccggatgg    11700
gccttcatgt agtcgagcgc ctggcgctgc gcccacttct ctttcgtgcc ctccgtccag    11760
cggctgccgt ccggcgcgcg aggcggcatc gtgcgcgacc attccttgtc gcccgtgagc    11820
gagatcgcgt cccacatgcg atcctcgggc gtgtacgcgt agttgcccat catcaggttc    11880
agcccgccga gcgtgtccga gaccgtgaac gagcgttgca ggagcgtgtt gcgcacgctc   11940
cacgggccga cgatccggcc gtatccggcg aagagaagca cggccacgcg cagacgcgcg    12000
cgaagcggga tgcccagacc gacgagcgcc gcggccatga ggatcgcgac gtacggccac    12060
atgatcgagc gaacgagcgc ggacgcgccg agcgcgacgc ccgtcgcgag cgcgatccac    12120
gccgagtgcc gccctgatgg gcgatcgatg agcgtcagac atcccaagag caacaaaaga    12180
aggagtgtga tgaagagcgt ctcggacagc accagcacgc cggagaacag cagcgaggga    12240
tagaacgcga acgccgccgc ggcaatgagg ccggcgcgtg cgctgaacaa cctgcgcccg    12300
atgagataca cgagccagac accgagcagg ctcaacggaa cctgcgccag ccggaccgcc    12360
gtcagcgatt ccgagcccgt gaccgcccag acgcccgcga cgaacgcggg aaagagcggc    12420
gcccggatgg acgtcggttc ccccggcccc cacgcgaacc cgttcccggc gacgacgttt   12480
ctggcaagcg tcgcgtagtg ctgctcgtcg cgaatctcga gcgcgacgtc cttgaacccc   12540
caaatcaggc cgagccggag gaccagggcc agcgtgagga tgacgagaat tgccccggcgc   12600
tcccaatggg tgcgctcggc ctcgtccgcc cggggggcga cgattggcca gggggccgac    12660
gatgggagtg ggggagcctg catcattcgc taaaacaagc agttacggtc aaacctgtgc    12720
caccggggaa tcagcaaaga ctctgccaac gggtccgccg gcggcgacgc ccggcgcgcg    12780
ttttccggcc ggaagtcggc ccccaaggcg acccggctgg gcacagccca gtagaaatct    12840
cattctggga tgactttgt ccatgatcca tgaccatttc gaggggcctt cgtgacgccc     12900
tggggatggc ataagcttgg cacctgcgat ggcgacgtgt cgatgatcag cggccgtgag    12960
gagctccaga aggcctatcg ggacgaccgg gtcgcccgtg aatacgtggc gcggcggttc    13020
cagtcaccgc tgggggcgct cctccattcg cgtcagatcg gggtcgtgcg agagctcgtc   13080
cgcgcgcagg gcatccggcg cgcggcggag atcgcacccg gtccagcgcg gctgaccgtg    13140
```

```
gacatcgcgc cgctcctcga ccgcgtgacg ctcgtcgacg cgagcgccca gatgcttcac    13200
gaagcgcgcg agcggctgcg cgaccggacg ctcgccgcgc gggccaatct cgtgcaggcc    13260
gacgcgtttc gtctgccgct cagcgggccg ttcgatctcg tctacacctt caggctcgtg    13320
cgccatttcg agcgcgccga tcggctgcga ctctaccgcc agattgcgac gattctcgga    13380
cccggcggct ggctcgtctt cgatgcggtg aatgagattg tgtctgcgcc gcttcgcgcc    13440
aacgcgaagc ccggcgagta cgagcactac gatgcgctgc tcacaccggc cgggctacgc    13500
gacgagctgg ccgaggccgg gttcacgatg gtgtcgctca ccggtgcgca tcgccggttc    13560
cccacgctca tgaagtgcca gatgtatctc gcgccgaggt caacggtgct cgcgcgcgcg    13620
gcgatggagg tgatcgatcg gctcggcggc gaaccgctcg agtggatcgt cgtatgccgc    13680
cgcgggtgac ctactggacg ggtacctggg atccgatcaa agaagcgatt tcgaaggaag    13740
tgaacgccct ccgcaccggg tcgcgggcga acgcgacggt cgtgtccggccgc gcctggatcg    13800
agcgcagccg tctcgtgccg aaagatcgcg tgctcgtgtt gtccggccgc gcctggatcg    13860
cccttcgggc cgcggccgcc gtgctcgagc cgcgcggcga cgtgacgcac gtcttcggcg    13920
gcgggttctc ctggcatctc ctgagggcgc tggggcgatcgc gcacgtggcc gtcgaagtcg    13980
tcaaggaaga tggcgctgcc gatctgcccg gggcgatcgc gtatgtctcc cggccgcgtg cacctcattc    14040
gccgcgctcg cgacacgtgg atcgcggccg gtatgtctcc cggccgcgtg cacctcattc    14100
atcccggcgt ggacctcgac tggtatcggc cggtgccggt ctcgtcggag agattcacgc    14160
tgctcttcgc cagcacgccg tccgatccgg ctgagatcga gccgcgcggc attccgctcg    14220
tcgtcgagct ggcgcggctc cgtccggaga tcgacgtgct ggtgccgtgg cgccggtggg    14280
gcgacgtcga cgcctccagg cgcgcgctcg acgccctccg tccgccaccg aacttcatcg    14340
tctcgtacga ggacgcaccc gacatgcgcg cctgttacgc gcgggcgcac gccaccgtcg    14400
tgtgcttcgc gggcggcacc ggcaaggcct gtccgaattt cgtgctcgag ggattggcca    14460
tgggacggcc gtgtctctcg acgcgcgaga acggcatcgc ggagatcctc gatcgatcgg    14520
gcgccggcat catcgccgcg cgcgatgccg cgtcgctcgc cgcgagcctg gacacgctcc    14580
gcgccggctg ggccggctgc gccgagcgcg cgcgccggct cgcagaggag acgttcgacc    14640
tcggccgctt ccgcgccgcg tacgacgcgc tgtacgtgga ggtggcatcc gccgccgctg    14700
gcgagcacgc gaggcgcgcc tcttgacgaa gggactcgcc ggacggctgc tcgggtgggc    14760
tcagaaggcg aacgatttcg ccgactacgt ccgccatcga tcgcagatga agcgcgacgc    14820
gcgggcgaat cgcccgcaggc tccatcgggc gctcggcggc cgcgtgaacc ggatcctcat    14880
cgtgcggctg gccagcatcg gcgacgtcgc gcgcgcgacg ggcgtcctcg gggccctgcg    14940
cgcgcagtac ccgcaggcga cgatcgactt cctgacgtcg gacgcgaccg tggccgtcat    15000
ccggcaccac cccgcgatca acaccgtcta cacgctccag gatctgtcgc gcctgtccga    15060
gtacgactgg atcatcaacc tccagaacct cgtctcgagc cgcagccgtt ccaggatcat    15120
cgggctctcg tacccgcaga ttctcgagca gctctcgagc acgaacatcc tctactgcat    15180
caccggccgt cacctcgaga acggctcgga gtccacgccg acgaacatcc tctactgcat    15240
ctgtgagatc gaggagcatt tcctcacggc gctcgtcccg tacgacacgg cccgctaccc    15300
ggtgaccgcg attcacgtgg acgcacaggc acgagcggcg gcacgatcga agttcgcgtt    15360
cccggagaac cgccaggtgc tggcgatctt cctcgggtcc aactcggtcg ggtgcggcgc    15420
cgatgagggc tttcgcacgt actccatcga ctacgtgaag cgcctcgtgc gtcacttcga    15480
agatcggttt gcgatcgcca tcatcgggca gtcccaggtg cgcaacccgg ccgaacgcgc    15540
acagtaccgc gagatgctcg gccgccatcc ggacgtgatc gatctcgtgg acaagacgtc    15600
gctcgacgag ctcgtggccc tcatggactc gttttccctg gggctgtacg tcaacgacgg    15660
ccccgtccat ctggcgatgg cccgcggagt gccggtcatc gggctgtacg tcaacgacgg    15720
gacgttcagg atgcacccga cgctcggcga cgagcgcttc gtggccatca acagcacgcc    15780
gccgtgcttc acctactcgt ggcgatggaa gttcttctgc tcgacgtgcc gcgacccggc    15840
gaccagatcg cattactgtc acaacgaggc gttcgtgttt ggcgtggacc gcattccgct    15900
ggccgcaatc gaccgcgcgg tcgatcgaga cgaagcgcaa agcgctctcg ctcgtcgatg    15960
ttcgcgcgca gtttgaagag ctcctcatcg aggcgctcga ggccggccga gctcgtgggcc    16020
cggtggccga gtcggcgcgc ctcctcatcg aggcgctcga ggccggccgag ctcgtgggcc    16080
tgtgcggcaa cggcggctcg gccgccgact cgcagcattt cgccgccgag ctcgtgggcc    16140
gcttcgagaa ggagcgccgc gcgttgcctg ccgtggcgct cacgaccgac acgtcgatcc    16200
tcaccgcgtg gccgaacgac tacgcgtacg agaccgtgtc agccggcagg tgagtgcctc    16260
ggagcgccgg gcgacgtgct catcggcatc atgaaacgaa tcgcgctcac cggacgcggc    16320
gccgccatgg acgaggcgga aaagaaacgc tgccaggtca tcgtcccgtc agatcagacg    16380
ggcggcaaga tgaagagtct cgccggcgtc tgccaggtca tcgtcccgtc agatcagacg    16440
tcgaggatcc aggaagtaca cattaccgtc atccacatct gggcgaagtt gatcgagaaa    16500
gcgctcagct gagctggccg tagacgtcga acgtgcgatc ttcgaacgtc gaccaggagt    16560
gctcgcgggc gatgaggtcg gcggcggcgt cggtgaggcg gcggcggagg tgctgatcct    16620
cgaacaacag ttgcagtcga ttcgcgtagc cctcggcgtc gtccgcctgg acgagaaagc    16680
cggtcttccc gtcttccacg acttcacgca cgcccgggat ggtggagccg atcacgggca    16740
cgcgggccga ctgggcttcg agcaccgaga gcggcagccc ctcgtgcgtt gacggcagca    16800
ggaaggcgtc ggcggcctgg agcagcaccg gcgcgcctcag cgcctgctcg tgttccttgc    16860
cgcgatcctc gagcgcgagc tcggcagaga ggcgcctcag cgcctgctcg tgttccttgc    16920
cctcggtccg gtcgctcccg accagccagc actcgaccgg caggccgcgc gacatcagca    16980
gcctggtcgc ccggattgcc gtagcctgcc ctttgtgtgg cgcaagattc gcgagcgcgg    17040
ggacgacgaa gcggtcggtc tcgaggccgg cggccgtacg cgcggcttgc cggtcgccgg    17100
ccacgaactt ctcccgatcg accgcgttcg gaatcgcggt gaccggcacc ttgtcccggc    17160
gtccctgctg cgccgcgtcg atcatcacgg cgatgtagcg cgcacacgtg acgatgtgct    17220
gcggcgggct cttcaggacc caggcgatct cctcgggatc cggttcgatg tgaaagtgca    17280
```

```
cggccacgcg cgcgcgcgcg gcctgcagcg cgggtctcat gaatccgtac acgacaggat   17340
tgtgcacgtg gaccagggggc cgcccgacga gacccaatcc cggaagcatg cgcgcacacg   17400
cggcgagctg ttgcaggacga ccggcccgca gcgcgttcag gtcatacatc cgccgccgga   17460
cgccctccgc gtcgagcgcg gcgcccgcag gaccctcgcc tgtaatccat gccgtgcacg   17520
ggacgccgcg gcgtttggcg gcggccgaca accggatcgc cagcaccgcc gcgccgccga   17580
ccagatgtga gacgaggatc tcgtgcaggc gggggggactg attcatggtg gcgcgcgatg   17640
cggccgggcc gccagcctac gttactcgcg tcgcaggtcc gcccgcgcgg tagagcgagc   17700
gcagcaggct gcgaagtccc gacgggagga agcacacggc gtaccgcgtg agatcggtgc   17760
gccgcatcgg caggcccgcc tgaccgaacc gtttccgcgc cgcgtgaaac tcgccgcggt   17820
cgtactcgag gtggccgagc gcgagcagga gacgatggcg atgatcgctg cgggcgcgat   17880
gcgccgactg atcgctcccc cgcgccagga accgctccag cgtcctgacc atgccttcga   17940
gcatcacgat ggacgacttc gacatgttgg tgtcgtgcag cctgtaatgc gtcaggggtt   18000
ccgggatgaa accgaccggg cccttcgccg agaggcgcag ccagaggtcc cagtcctcga   18060
cgccgtggat gccggtcgcg aacatctcct gccggagcac ggcggctttg accatgaccg   18120
acgaagcggt cacggcgttt cgcaggacga ggtcggggag acagcgcccc tcgagcggcg   18180
gaagatcgag cgtgcgcagg acctcgccgg cctcgtcgat gacggcgcgg cccgtatgaa   18240
tgaagctcca gtcggcgtgc tgtcggaaga gatcgacctg ccggcgaatc ttggccggat   18300
cccacgagtc gtcggcgtcg agcagcgcca cgaattcccc tcgcgccgcg gcgatggcgc   18360
gattgcgggc gccggcaagg ccccggttgg cctgagagat caccgtgatg gcgccgccga   18420
atctcgccag gatgtctgcc gtgccgtcgg tcgatccgtc gtcgacgacg atcacttcga   18480
cgggctgata ggtctgggag agcgcggact gcacggccct ggcgatgtag cgctcgccat   18540
tgaacaccgg gacgatcacc gacacgaggc cggggatatc ggtctgggcg gccggcgagg   18600
gctctcgagc ggcttcagcg gccatggccc tctgcactgg cgtcgcgagc gaacagcatg   18660
acgtttctcg aaagccacgg agacaacggc gcgctcttca tccgctcccg cagcgtctgc   18720
gatgccgggc cgaggtagct gaacccgcgt gccgcgaaga tcttctgcca gtactcgagc   18780
ggctgttcgt tgatgtggcc cattccgccc tggcccggct gcgcggcgct gaagagcacc   18840
agcggcgccg tgttcgtgag cacggcgacg agcttctcgc caagcggcgg cggaaggtgt   18900
tcggcgacct cgaggctcat ggagaggtcc gttgcgtccg cggccgtgga caccgccgtg   18960
agatcgaatg agttcagcgc gacgccgagc ttttccgcgg cgtaccgtcg cgcgtactcg   19020
gatcgctcgt agccggccgc cttcacgcct cgtgcgagga acgcgtgcac gaagtgaccg   19080
gtgccgcagc cgaagtcgtt catcgatcgg atgccgggtc gcaacgccag gaccccatca   19140
acgagcacgg gcgccgcctc ggcggcgacg tgcccgaaca gatccatgac ggccgcgttg   19200
tacgtccatc gctcgaagcc cagccgttcg ccgatccggg acatcttcag tccggcctgg   19260
caacgccccg accagtagag ccgcttgagc tgggtcacga tcatgggttc gtcccgggcg   19320
cgacgcgaac gcgcggcggc gcatgcagcg acgccgagaa ctccacatcg gccacgccgg   19380
cccacgatcg atgttccttg accgccagcg tcgcaagctg ccgcgacggc acgaggaacg   19440
tctgggtcgg gttgtgaaac acgtgcaccg acaggtagta gtggccgcga acgagatgcg   19500
cgcggaaatg gaggtcgcag gagaactcgc ccatgaacgc ggacgcgtcc ccgagatccc   19560
gccggtggat gttgccgtcg tagacgacga gctgatcggt ggagcgttgc accatgaagc   19620
cgaacgtgac gtcgtccagc ggctcggtac ttctcgcggt gacctggagg ttcagctgct   19680
tccccggcgt cacttcgctg tgcgtgcccg acgactcgtc gccgacgaag agcgagccgg   19740
acacgatctc gacggggctc ccgacactcg gctgttccga tttcctggac gcgcggacgt   19800
acgcgtcgag gacttcgatg gccgtgccgg cggccttcac cgagccggcc agatacaccg   19860
catgctcgca gagggtcgcg accgcctgca ggttgtgcga gacgaggacg atcgcgatgc   19920
cctggcgctt gaactcctgc atgcgatcga agcacttctg ctggaacgtc atgtcgccga   19980
cgctcaacac ctcgtcgatg atcagcacgt cgggccgcag gtgcgcggcg atcgagaagc   20040
cgagccgcgc gttcatgccc gacgagtagc gcttgacggg cgtgtccacg aagtcggtga   20100
cgccggcaaa ctcgacgatc tcgtcgaacc gctcggcgat ctcgacccgt ttcatcccca   20160
tgatcgcacc ctggaggaac acgttctcgc gcccggtgag gtcgggatgg aaccccgcgg   20220
cgacttcgat gagcgcgacg gcgcgccccc ggacgatcga gcggccggtc gtcggcttca   20280
ggatcttcgt cagcaccttg agggtggtcg acttgcccgc cccgtttgcg ccgatgatgc   20340
cgagcgcctc gccgcggtgg acttcgaacg agacgtttct gaccgcccag aactcctgtt   20400
cgtccagcgc gtcggcggga cgcggttgaa acagcccgcg cacgagcgcc ggcacgaggt   20460
cccgcaggct gtcatgacgc tcgccccgcc ggaacttctt ggagacgttg tcgaagacga   20520
cgggggccgt cacgtcagac gttctccgcg aattcgaact cggagcgatg gaagaacagc   20580
cagccgaaga cgagcgtcac gaccgagacg atcgccgtga cgcccagcat cacggacggc   20640
ggctggttca ggagcagcac ggcgcggaag ccgtcgatga tcggcgtcat cgggttgagc   20700
tgcagggccc actggagctt gccgcccacc atttcgaccg ggtagaccgt cgacgacgcg   20760
aacatccaga cggcgatgac gacctcgaag aggtatttca cgtcgcggta gaagaggttc   20820
gccatcgaca ggatcagggc gatggccgtg gtgaatatca cgagcacggc gagcacggcc   20880
ggcagccaga ggatgttcgg tccgaccggc acgtggtagt agatcatgag cgcgacgagc   20940
accacggagc cgatcgcgaa gtccacgacg ctcacgatca cggcggagaa cgggaagatc   21000
tccctgggaa agtacacctt cgagacgaga ttcgtgttcc cggtcaggga ggtcacggag   21060
aagcggagcg ccgaggagaa gaagttccag acccacaggc cgcagaacgc gaacaccgga   21120
tacgcgaccg gcgtgtcgat ggccgcgacg cgcatgaaga tcacggagaa gacagccgtg   21180
ttgatgagcg gcatgaacac ggcccacccg aagcccatca ccgactgctt gtagcgcagc   21240
agcaggtcgc gccgcgtcat ctgcatcagc agctcgcgat actcgaactg ctcggcgagc   21300
atctcacggc aatcggccag gaaggcgcgc atcgtcagcg gtgtcagtcg cggcgtcgtg   21360
ccgggcggat catgagggca ggtagcgctc gagctgcccg aagagcaact gcacgaactc   21420
```

```
ggtcgcgcgc cgctcggcgg cgacgtcgcc gttttcgctc gacggcacgg ggctcgtcac   21480
gcggacgagc gcggcgtccg atcggttcag ccgcatcgcg tcgtacacca tgtagatctt   21540
gctccagtac tcgctggcga tcacgcggcc gtggctctgg taccagtaca gcacgatttg   21600
ccgctcgatc cccttctgga tgacgtatcg gttgatctcg atgggctcgg tgcggcccgg   21660
caccttgacc agcgcgcggc cggtggatac gggctcccag ccggcgcccg gcaggcagtt   21720
catcggcgag tggatcgtgt cgccttcgcg ctggctctgg tagtagccga tgtagagcga   21780
cacgtagggc tcgcgcggcg ccaggtagat ccggttgatg tactcgtcca cgccgagcac   21840
cgccatgacg tccttggtga agggcgcgct gtcgcgtccg gtccaccgct ccatctggaa   21900
cgggacgttc gcgagcggct ggcgaagcgg gacctgctcg cgcgcatcg agcgcgcggc    21960
aaacgccgcc gtgccgagga aacacacgct gagaatgatc agtcgtctgg tcatgaatct    22020
ggactcacac cgtggccggc gcggtcaccg ctcggggcgg aggcgtgaag cgccgcagga   22080
cccagaccac gacgaacaac atgacgaacg cggcaagaaa gacgagccag cccgagaacg   22140
tgtggaagaa cccctgcgcc gcggcctcgc cgtagtagtg cgccgccacg cccgtgccgg   22200
ccacacgcgc gccgttcgcg ataatcgcga tcgggatcgt cgagagcgca atcgccaggc   22260
gcatgccgac gcgcggatcc gtgaagtagc cgtacacgat gcccagcgtg agcagcgaga   22320
tgagcgaccg gatcccgctg cacgcctcgg ccacttcgag cgtcgtgtgc gcgagcacga   22380
tcacgttgcc ttcgcgcagc accgggatgc tgagcgcgga cagcacgccc tcgcccacgc   22440
gagacgccag gagctgcagc ggaaatgcga tctgattgaa gatgatcgac gggatcggga   22500
tcatcagcac caggaacgcg agcggaaacg cgagcaccca gagcttgcgc cagccgagca   22560
ggaacacgat cgcgccggta agcgaggcta gtatcgacac gcgcgtcaga aagagctcgg   22620
cgccgagcaa ccccgcgacg agcagcgcga gcccggcgag cacgatgacg agaccgagcg   22680
cgctcggcgt gtccggcagc gccgagagac gcgcccggcg ctcccaggca aagtaggccg   22740
cgagcggtac gatcagaatc ccgtgggagt agttgtcgtc gttgatccag tcctgcgcca   22800
accgcacgat cacctgccca taaaggagca ggaaactcac ggcggcgata ccgagcggaa   22860
ccagtctggc ctgaagcaca ctcatagcga tcgggacgaa tatagcaaag ggctctcccg   22920
ggcagaactt gggcgccgaa acacgtaacc atcacgctca caggaagtta gcggctaaat   22980
gacgggattg gccggctgga ggggcctcgt gaatctaccc aatggatgac aaatatccct   23040
atcaactcct ccttttgact gcgtagcacg gaagatcatc ggcctcttca gatgcaggac   23100
aaatgtcatc tgtcaggaag agtaatgtct ctggccacag cgccggtcaa gtgaaatccc   23160
ccgacagaat cgcgtaccac cgcgactggc atccggcgag aaaagtcgcc tcgtcggccc   23220
gaaaactgca ttacaggaac gccaggtcgg cagccggttc gttttccgtc agtattttg    23280
ggcattgcag gcgctttcgt ctgccggggg gcggaaccgt gcgaccgggc cgatcacgcg   23340
tggagcaggg cagggacagg catggatggt tccgtgagac agccgcgttc gggagacggc   23400
cgggctcggc gggagctcgt ccacctggcc gtggaccgcg cggcgcgcca atttcctgac   23460
gcgccggccg taagttttga cggcgtctcg ctctcgtatc gcggcctgac tcgccgcgcc   23520
aacgcgctcg cgcaccgcct gcagcgtcat ggcgtcgatc ctgacgtggt ggtcggcatc   23580
tccctcgacc ggtccatcga gctcatcgtc gggatcctcg ccgtgctcaa ggccggcgag   23640
gcctacgcgc cgctcgatcc tgcgtatccg aaggaccgcc tcgagtacat ggtcagggac   23700
gccggggcgc ggatcgtgct cacctcgcga cgcgtgagcg gtctgatccg cgtcgatggg   23760
gtcgagctca tctgcctcga tgcggacgag cccgggggcgt tggcggaaga gggccatgcg   23820
ccggacgccg gcgtcgatct cgaccacctc gcgtacgtca tctacacctc gggatcgacc   23880
gggcgcccga agggcgtcgc gatgccgcat cgtccgttgg cgaatctcat cggctggcag   23940
gtcgacgcat cggccgtcgg cgccggcgcg cgcacgctcc agttcacttc gccgagcttc   24000
gacgtctcgt tccaggaaat ctttgcgacg ctcgcgaccg gcggcgagct cgtgctcgtg   24060
tccgaggaca cgcgccggga tccgcgcgcg ctgctgcggg tgctgcgcga tcgatccgtc   24120
gcccggctgt atctcccgtt cgtcgcgttg cagcagcttg ccgtcaccgg caccgacgat   24180
ccggatctgc cggcgttgcg tgaggtcatc acggccggcg agcagctccg ctcgacgccc   24240
gcgctcgtcc agtggttcgc gcggcatccg cagtgcacgc tccacaacca ctacggcccg   24300
gccgaagcgc atgtcgtcac tgcgcacacg ctgactagcc ccccgcggga atggccggcc   24360
ctgccgccga ttggcgtgcg gctgcccggc gtcgagattc tgatcgtcga cgagcatcgc   24420
cagcccgtcg cgcacggcaa cgtgggcgag atcctcattg gcggcgtgtg cctcgctcgc   24480
ggctacttga atcagccgca gcgcacggcc gaacggtttc tctcgcatcc gcttcgcccc   24540
gatgctcgcg tgtatcgcac cggcgatctc ggccgtgtgc gcgacgatgg cgcgatcgag   24600
ttcgtcgggc ggacggacga ccaggtgaag attcgaggct tccgtgtgga gccgggcgaa   24660
atcgagaccg tcctccttca gcatcccggt gttcgggaga ccgccgtcgt cgcgcacgac   24720
gacgcgtcgg gcagcaagag gctggtcgcg tatgtgattc cgtcagccga tcaggcggcc   24780
accggagccg aggcttcggc acaggtctcc ggatggcaga ccgtgtggga ggacacctac   24840
cgcgcggcgg cggactcgac ggatccggcg cttgccgcga ccggctggcg gagcagctac   24900
acggggatgc cctacaccgc ggcggagatg cgggagtggg ccgatgagac cgtcgcgcgc   24960
atccggtcgc atgcgccacg cgaggtgctc gagatcggct gcggcacggg catggtgctc   25020
ttccgtctcg cgccccactg cgctgtctat agcgcgcggg atgtctctcg cgccgcgctc   25080
gaccacgtcc gggcccacgc gacggcctgc ggcgccagcc acgtcgagct cctctcgcgc   25140
gaggcggcgg actttaccgg cgttccggcg ccgcgttcg atgccgttgt gatgaactcg    25200
gtcgcgcagc acttgcctgg cgtggactat ctgacgcgcg tcctcgccgg cgccgcccgc   25260
accgtcgtc cgggcggatt catcttcgtc ggcgacgtgc cgaacctgcg cctgctcgag    25320
atgtttcacg cgtcggtcga gctcacgaaa gcgccggcgg atcttccgat tgccctcgtg   25380
cgcgagcgcg tgcagcgtca gatggcgctc gagcgcgagc tcgtgatcga tccgatgttc   25440
ttcgacgccg tcatggaccg catccccggc atctcgggag cccaggtcca gatccggcag   25500
ggtgccttcg acacggagat gaaccggttc aggtacgacg tcgtgctcac ggtgggcgga   25560
```

```
cggccggccg ggcgccgtcc gtccgtcgcg ctcgactggc gtacggagcc ggtcgccgag    25620
cgtcagctcg ggcaactgct cagcgccgat gcgccggacg cgatcgagat caggaacgtg    25680
accaacgcgc ggctgcgtca gccggccgcg tgcttcgatc tcgtcaagcg tgccgactgg    25740
accggcacgg ccggcgagac gcttgagagg gccgccgcgg acacgcgcga cgcgatcgac    25800
ccggaatcgt ggcggcgtct cgcgaacgcc gccggctacg acgccgagtt cacctggtcg    25860
gccgagatcg gcgccatcgg cttcgatgcg cggttcgagc gccggggcac cgcgtcgctc    25920
gccgctcccg tgactcgtca cgcggttccc gccgcggagc cgcgggccat tcagcagtac    25980
gcgacgaatc cgctccgcga tgcggtggcg cggcgcctca cgcccgagct tcggcgattc    26040
ctgcaggagc gcgtgccgga tcacctggtg ccgtcggcct tcgtcgtact cgatgcgatg    26100
ccgctcacgc caagcgggaa gatcgatcgg cgcgcccttc cggcgcccga gagccgccgt    26160
cccgatctgg acgtggcgtt tgccaagccc gccacggagc tcgaacgcaa gatcgcgcag    26220
gtctggcaga cgacgctcca gatctcgtcg gtgggattgc acgacaactt cttcgatctg    26280
ggcgggcact cgctcctgct cgcgcaggcg ttcgagcgta ttcggccgct cgtgccgggc    26340
aagcagtggt cgatgatcga gatgttccag tacccgacgc tctattcgct cgccgggttt    26400
ctcagtgagc ggccggacgc cggcgccggg cagcttgccg gccggcagga tcgcggacgc    26460
cgccagcgcg agatgctgac gcgtcagggg cctgccctga gcagagtcga aggggtccgg    26520
taatgcagaa cggcatcgcc ctcgtcggga tggccggccg ctttccgggc gcgcggtcga    26580
tcgaggagta ctggcgcaac ctcgtcgccg gcacggagtc gatcacggtc ttcaccaacg    26640
aggagctggc gacggccggg atcagccaga gcgatctgca gaatcccgac tacgtccgcg    26700
cgcgcgggct cgtcccggat ccggacctct tcgacgccgc gttcttcggc atcagccagc    26760
gcgaagccga gctgatggat ccgcagcacc ggctcttcct cgaggcgtgc tggctcgcgc    26820
tcgaacacgc ggcgatcgcg ccgcggtcgt tccccggtct catcggcgtg tggggcggca    26880
tgagcacggg gatgacgaac agcacgtacc tgctgtcgaa tctccacagc cacgcgggat    26940
tgctccagcc ggaggacgtg ctgccggcgc tgctcggcaa cgagaacgac tacctcacga    27000
gccgcgtctc gtacaagctg cacctgcgcg gaccgagcat gaacgtgcag accgcgtgct    27060
cgacgtcgct cgtcgcggtg gtgcaggcgt gtcagtcgct gctcacgtgg cagtgtgatg    27120
cggcgctcgc cggcggcgtg tccgtgtcgt tcccgcaaca ggaaggctac gtctacgtcg    27180
agggcggcat cgggtcgcca gacgggcact gccgcgcgtt cgatgcgaac gcccaaggca    27240
cggtcttcag caacggcgtc ggcatcgtcg tcctgcgaag gctcgaggac gcgatcgccg   ,27300
acgggcagac gatctacgcg gtgatccgcg gatcggcgct gaacaacgac ggctcggcga    27360
aggtcagttt cgcagccccg agcgtggacg gccaggcgga ggtcgtggcc atggcgcaga    27420
gcgccgccga cgtggcgccg cagacgatcg actacgtcga ggcgcacggc accggcacgc    27480
cgctcggaga tccgatcgag gtggctgcgc tcacgaaggc ctttcgcgcc ggcggcgcca    27540
cggagaacgg gttctgcgga ctcggatcgg tcaagagcaa tttcgggcac ctcgactcgg    27600
cggccggcgt cgcggggctg atcaaaacgg cgctcgcgct ccatcaccag cagattccgg    27660
cgacgctgca ctacgtcgag cccaacccga agtgtgattt cgccaacagt ccgttcttcg    27720
tgaacgccgc gctgcgacca tggccgcgcc gtggccatcc gcgacgcgcg gccgtgagct    27780
cgttcgggat cggcgggacc aacgcgcacg tcgtgctcga ggaagcgccg cccgaccggg    27840
agcgaccggg cgccgagccc gccagcgacc tgctcgtgat ctcggcgaag acgccggcgg    27900
cactcgacgc ggcgcgctgc aacctcgcga gtcatctcga ggcgcacccg gacgtcagca    27960
tcgccgacgt ggcgcacacg ctgcgcgccg gccgacagga gtttccgcat cgacgagcgc    28020
tcgcctgcac cagccgcacc gatgcgatcg ccgtgttgcg cggcgcggac gccaaacgtg    28080
tcgccaccgg cgacgccgaa tccggcagcc gtgccgtggc cttcatgttc tccggcggcg    28140
gcacgcagta cgcgaacatg ggccgggagt tgctgccggt cgcgcaggtc ttccgcgaag    28200
aaaccgagcg ctgcctcacg cacctgtcac tctccgccga cgtccgccgt gcgctgtttc    28260
cggatcatgc ggacgtcgag cgggccgcgc attggctcga agctccgtcc gttggtttgc    28320
ccgcgctctt tgtcgtgcag tacgcgctcg cccgtcagtg gatggcgtgg ggtgtgcagc    28380
ccgacgggct catcggtcac agcatgggcg agtacacggc cgcgtgcctt gcgggcgtga    28440
tgtcgctcga gaacgcgctc gcgctcgtgt gtctgagagg ccggctgttc gagacgcttc    28500
cggccgggga gatgctcggc gtcccgcttc ccgaggcaga ggtgcggccc gcgctcggcg    28560
atggactcgc gatcgccgcg gtgaatcgc caagctcgtg cgtcgtttcc ggcgcgcccg    28620
acgcgctcgc ggcctttgcg cgcgtgatgg ccgaggccgg tgtcgaatgc cgccgcctcc    28680
acatcgcggt cgcggcgcac tcgccgatgg tcgatccgat cctggcggag ttcgagcgtt    28740
tcctccgcgg catcgcgctc gacgcgccaa gggttccagt cgtgtcgaac ctcagcggct    28800
cgtggctgac cgcggacgaa gcgaggtcgc cgcagtactg gacgcggcat ctgagacaaa    28860
ccgtcaggtt cttcgacggg ctcggcacgc tgctcgccga gccgaatcgg gtgctcctcg    28920
aagtcgggcc cggacagacg ctgtccggat tcgcacggca gcatccacag aaaacggcgg    28980
cgcacgtcgt tcaagcgacg ctccggcacg ccaaggaacc ggcgccggac gtcttgttcc    29040
tgcagcaggc ggtgggcagg ctctggacgg caggggtgcc cgtgaactgg gctgcttaca    29100
acagaggcga gggcaggcgt cgtgtgccgc ttccggccta tccgttcgag aagaagcggt    29160
acctggtcga atcgccccag cgcgccggtcc cagcgccgct tcaagccctg gcgccaaggg    29220
ccgatgctac tcagcagatg acgacaactg tgccttcgtc tccggcccga tcgcgtgccg    29280
atcgtatccg cgagcggctg accgaaatcc tgcacaaact gagcggcctc ccggcaggcg    29340
agatcgatcc cgcgctgtcg tttctggaga tgggattcga ctcgttgttc ctgacgcagg    29400
cgagtctgcg gttcaaggcg gagttcaagg tccggatcac gttccgccag ctcttcgaag    29460
aggcaccgac gatcgacgcg ctcacgagct acatcgacgc caagctgcct gcggacgcgt    29520
ttcctgcgcc ggcggaagcc gctccggcgc ctttcgtggc cgcaagcgtg ctgtccgcgg    29580
gccctcatgc ggccctggct ggcgtgccgg cggccgcgcc cggcacgctg cagcacgtga    29640
ttcaacagca attgcagttg atggccgagc aactccgcat gctgggcgga aatcccgccc    29700
```

```
tgctgcccgc gatgccgccg ccagcgtcgt ccgtcgcggc ggccgacgcg ggactgaagt    29760
cccgcgctac gacggcgccc gtctcgaagc cgcatgcggc cttcaagccg atcgacggca    29820
aagccagtgc ggagctgtcg cccgagcaga tggccgcgct cgccgcctgg atcgatcgat    29880
acacacggcg cacggccgga tcgaagcgac gcctggccga gtaccggccg gtgcttgccg    29940
atccacgctc ggtcgccgga ttcaaccggc tctggaagga gatgatctac ccgatcatca    30000
ccgaccggtc ggaaggatcg aaggtctggg acgtcgacgg caacgagtac gtggatctcg    30060
tgggcgggtt cggctcgctt ctgttcggac accgttcgcc gatcgtggtc gacgccgtgg    30120
aggaacaact ccaccggggg ttcgagctcg gaccgttgcc gccgctcgcc ggcgaggtcg    30180
cggcgctcgt caaggaattc acggggatgg agcgcgttgg attctccaac accggctccg    30240
aagccgtgct ggcggccacg cgttttgccc gaacggtcac cgggcgcgac aagatcgccg    30300
tcttcgaagg gtcgtatcac ggtctccttg acgaagtgct cagccggccg ctcgtggtga    30360
acggcgaatt gcggtcggcc cccgcggcgc ccggcatcgc cggcagcgcg gtgagcgaag    30420
tcatcctgct cgagtacgcg aacccgaagt cgctcgaggt catccgcacg cacgagtcgg    30480
agatcgccgc gatcctcgtc gagcccgtgc agagccggcg gctcgacctg cagccgaagg    30540
agttcctgca cgaggttcgc cgcgtggcgg acagagatcgg cgccgcgctg atcttcgacg    30600
aagtcatcac cggcttccgc gtgcatcccg gcggcgccca ggcgcacttc ggcgtccgcg    30660
cggatcttgc cgcctacggc aaggtcgtcg gcggcggcat gcccatcggc atcgtcgcgg    30720
gcaatgcgaa gttcatggac gcgttcgacg gcggccactg ggagtacggc gacgggtcgt    30780
tcccggaggt gggggtcacg tactttgccg gcacgttcat gcggcatccg ctcgcgctcg    30840
ccgctgcgaa ggccgtgctg acgtacatga agtcgcaggg tcccggcctg cagacgcggc    30900
tggcagatcg gatcgagcgg ctcgcgaacg acgcgcgcgc gatcgtggcc aggcacggcg    30960
cgccgtatca gattacgcag ttcagctcga tgatgagcct gaacttcccg cacgaccaga    31020
agacggcggt gctgctgttc ttcctgatgc gcgagcgcgg gattcatatc tgggagggcc    31080
gtcccttctt cttcacgacg gctcacaccg aggccgacta cgccgcgcgcc ctccgcgcgc    31140
tcgacgagtc gctggccgag atgcaggcgg ctggcttctt cggcgcgaac ggacagcagg    31200
cattgcaggt gaccgggacg gcggaccagg acgtcctgcc gttcaccgaa ggacagcagg    31260
agatctggct cggcggccag atgggcgagg cggcatcacg cgcctacaac gaagtcgtcg    31320
cgctcgacct gcgcggggcct ctgaatcgcg cggcgatgca gcgcgccgtc gacgaggtcg    31380
tcgcgcgtca cgaatcgctg cgcatgacgg tggccagcga accgctcggc ctgcgtcacg    31440
ttccgggcac gtcggtgccg gttgaatggc aggacgtgtc gggattgccg gaggacgctc    31500
agcacgaggc cgtgcagaag gtgctcgagg cggcgacgg cgtggcgttt gacgtggaac    31560
gcgggccgtt gctgcgcgtg accaccatga cgctcacccc cgagcatcac gtcttcgtca    31620
tggcggccca ccatctcgtg tgtgacggct ggtcgttcgg cgtcatcctg cgcgatctgg    31680
cgacgttcta ttcggcccag gttcgaggca cgcgtgcgaa cctcgaggcg ccgatgcagg    31740
tgagccgctt cgcgcgtgag gatcgcgagg ccaagcagag tcctgaagcg gccgagaccg    31800
aggcgttctg gatccggcaa ttcgacaccg tccccgagcc gctcgagtta cccgccgatc    31860
ggccgcggcc gccagtgaag tcgtaccagg gcgcccgtgt gtccgtgccg tttgacgcgg    31920
cgctcgctcg cggcgccaag aagctcgccg ccgaacaccg cacgaccctg tttacgacgc    31980
tgctcggcgt cttcgagacg ctcgtctacc gactgaccgg ccagcaggat ttcgtcgtcg    32040
gcatcccggt cgcggtgcag cctctgctcg gcgaggatct ggtcgcgcac tgcgtcaact    32100
tcgtgccgct tcgcgcgcgc gtgtcaccca ccgcgacgtt ctcggagtac ctgaacacgc    32160
tgaggacgca gtcactcgac ctcaatgagc accagaactt cacgtacggc agtttgctcg    32220
ccaagctgcg gttgccgaag cacccgggcc gcgatccgct cgtctcgacg agcttcacgc    32280
tcgagcccgc catcgtcgat cccgggttcg acggtctcac ggcgcgttcg ctgacggttc    32340
cgcgcgtgac atcgaagcgc gacctgcacg tgaacgtcat ggagatggac ggcggtctgc    32400
tggtcgaggc cgactacagc accgatctgt tcgacgaacc gacggtcgcg cggtggctgg    32460
atagctatcg gatcctgctc gaggccgtcg tgtcgtcccc cggccggtcc ctcctgagcg    32520
tgccggtgat ctcggagccg gaccggagtc agcttgtcac cggctggaac gacacggccg    32580
ccgactaccc gcgcgatcgc gtcatccatc aactcgttga agagcgcgcc gtgcagaccc    32640
cggagggact cgcgctcgtc tgtggtgccg agcggctgac ctggaccgag ctcaaccgac    32700
gcgcaaaccg gttggcgcat gcgctcgtga agaagggcgt cgcccccggg agcagggtgg    32760
cgctgtgcct ggaacgatcg gccgggttca tcgtctcggt gcttgccgtg cacaaggccg    32820
gcgccgcata tgtaccactc gaccccgtct cgcccggcga tcgcaagtcc ttcgtcgtcg    32880
aggattccgg tgccgtgctg gtcgtgaccg attcgcggtc cgaaggccga tggctgacgc    32940
ccaggacgcc cgtgctccac ctcgacgccg acagtccgcg aatcgcgcaa gagtcgcacg    33000
atgatctgaa ggtcgccctg tcggccgaag acgccgcgta cgtgatctac acgtccggat    33060
cgacgggaca accgaagggc gtcgtcatcg ggcaccgcca gctcgtgaat tacacgtggg    33120
cggtgatcga acggatcggg ttacccgccg gctcgagtta tgcgctcgtg tcgaacgtcg    33180
cggccgatag cggtgtgacc gggctctcgg cgctcgcgat gggctgggtg ctgcacgtga    33240
tcacctcaga ggtcgcgacc gacgggcggg cgctcgggtc gtatttgacc gttcacgaca    33300
tcgacttcct gaagatcacg ccgtcgctgc tcgcgtcgct gctcggcgat cacccgtcgg    33360
cctcgctgct gccgcgccgc tgtctcatgt tgggcggcga gccgtcacgt ccggcgtggg    33420
tggaagagct gcagcgtctc gcaccaggct gccggatcat gaaccactac ggcccgacgg    33480
agaccacggt cggcgtgttg acgtactgga ccggtgaccc cagggatctg ttcgtcgaga    33540
gcgtgccggc cggcaggccg ctcgggaacg tccgcgccta cgtgctcgat ccggccggag    33600
cgcccgcgcc gattggcgtc gtcggcgagc tctgtatcgg cggcgcgtgc gtggcacgcg    33660
ggtatctcaa tcgtcccgac ctcaccgccg agcgcttcgt ccccgatccg ttcgccaccg    33720
agccgggagc gcgcatgtac cggaccggcg accgcgcccg gttccgggcc gacgggaaca    33780
tcgagttcct cggtcgcgcg gactatcagg taaaagtacg cggattccgc gtcgagctcg    33840
```

```
gcgaaatcga agcggcgctt cgcgctcatg agggcgtcga gcaggccgta gtcgtgctgc    33900
gaaaagatca gcccggcgac gagcgcctga tcggctatgt cacgaccggc ggcggcgcct    33960
cgatctcgat ggcggagctt cgcacgtcgc tcaagcagag gttgccggac tacatggtgc    34020
cggcggtgat cgtcgcgctc gacaagctgc cggtcacgtc gattggcaag atcgatcgac    34080
gcgccctgcc ggcgccgccc gagcgcgcca gtttcgagtc ggagttcgtc acgcccgcca    34140
ccgagacgga gcgccggctt gcggagatct ggagcgcggt gctcggtctc gagaagattg    34200
gcgcgcacga caacttcttc gacctgggcg ggcactcgct gctcgcggcc cgcaccgcca    34260
tgaaggtccg cgacacgttt cgcgtggacg tgtcgctcat cacgttcttc gaaatgccca    34320
ccatcgccat gatggcggca gcaatcgacg ccggtaccgc ggccctgcc tcccggacca    34380
tcacgccgag ggcccgcgtc agggcccgca gggatgagct ggtccgccca tgaaccccgg    34440
ccacccgcag actgagtccc ccgacctgcg gcgcgacggg atcgccgccg aagcgggcga    34500
cgtgtttctg gcgccggcgt cgtacgccca ggggcgcttg tggttcctgg cgcagctcga    34560
gcctgaaagc tccgcctaca acattccgat cgcgctcaga ctgaagggcc gcctcgaccc    34620
cgccgcgctc gaacgcgcca tcaacaccat cgtccagcgt cacgagacgc tgcggacgac    34680
gttctcgctg cagggcgatg agctcaagca ggtcatctct ccgacgctgc gtattccgat    34740
ctccctcgtg gacatgcgtc tgctcagtcc tgagaaacgc gaggccggat ggcgccgcgt    34800
tgcggccgac gacggcgcgc gcgtcttcga tctggctgag ggcccgctgc tcaaggtcgt    34860
actcgtcatc ctcgagccct ccgatttcgt cctcctgatc accgtgcatc acatcgtctt    34920
cgacgggtgg tccggcggga tcttcatccg cgagctggcc gaggcgtacg cggctgagct    34980
cgacgggcgc gacccgaagc tgccggagct gccgatccag tacgcggact ttgccgcgtt    35040
ccagaaggaa cagaccgaag ggcccggcct gccggtgctg ttgaagcact ggacgcagga    35100
gctggcgggg gcgtcgatga cgctcgacct gccgacggat cgtccgcgtc cgccctgca    35160
gacctcccgg ggcgccctcg cctcgcttcg gctcgacgga gcgatctggc cgggcgtcgc    35220
cgcattcagc cggggcgaga acgcgacggt gttcatgacg cttctctccg ccttcggcgc    35280
cgtgctgcat cgctggtccg gacaggaaga catcctcatc ggcagcccgt cggccggccg    35340
cgatcgtccg gagcttgagg gcctgatcgg gttcttcatc aacacgttcg tcctgcggct    35400
caacgcgtc ggagatccca cgtttcgcga gttcgtcggc cgggcgcgcc gcgcgtgccg    35460
tgcggcctac gcgcatccag acctgccgtt cgatcggctc gtcgagacgc tcaacccgga    35520
gcggacccgg gatcggcatc cgatctatca agtcatgttc gcgcaggatc cgccggcgtc    35580
gaagatgacc ttcgccggga tcgagctcga gcggcttgtc gtccacaacc gcagcaccaa    35640
gtgcgacctc acgctcgagg ttgccgaaga cgccgacggc gtcaccctgt acgtcgagta    35700
cagcgtcgac ctgttcgatg cggcgacggt ggcgtcgctg ctgcagcaat atcgccgtgt    35760
gcttcagacc gcgatcgcgg acccggacca gcggctctcc gagctgcgtc tgcttggcga    35820
cgacgagcgc atgcagctga tcgcactggg gaccggtccg gcggccgtct acccgcgcga    35880
cgtcagtctg gccacggtct tcgaagagcg ggtccgcgcg actccggcgg cagtggccgc    35940
gacgctggag ggacagcacc tcacgtacgc ggatctcaac cggcaggcga accgcctggc    36000
ccggcggttg aaggcgttcg gtgtgggccc gaacgtgctg gtcggcgtgg cgctcgaccg    36060
gtcgttcgat ctgctggtgg cgctcctcgc cgtggccgaag gccggcggcg cgtatctccc    36120
gctcgatttc gctcacccgc aggaacggct tgcgttcatg ctcgccgaca cgcgcgcgcc    36180
catcgtgctc acgctgcggc gatttgccgg ggcgctcgag tcctttcccc tcttctgcct    36240
ggacgacgag ctgccggcca ccgccgagga gccggacgaa gatctgccgc cgcagtcgac    36300
cggagaccac ctggcctatg tgatgtacac ctcgggctcg actggtcagc ccaagggcgt    36360
cgcgatcacg catcgcggcg tggttcgtct cgtacgcggc acagactacg tccgctggaa    36420
cgacgtgcgg gcgatcctcc acatcgcgcc gacgtcgttc gacgcgttga cgttcgaggt    36480
gtggggcgcg ttgctgaacg gcgcgcggct ggtgctggtt cccgagcggc tcgtcagcct    36540
ggagacgctc gagcgcacgc tccggtcgga gcaggtggac tgcctgctgc tcacgaccgc    36600
gctcttcaac gccgtcatcg acgcgaagcc cggcattctc gtcggcgtga agcaactcct    36660
gatcggcggc gaagcgctgt cagtcgcgca cgtgcggcgc gccctggcgc acctgccgga    36720
cacgcggctc gtgaacgcgt acggtccgac cgagtgcacg accatcgcgt gcgcctacca    36780
gattccgcgc acgctcgatc cgcacgcgcg gtccatttcg atcgccgacg cgattgccaa    36840
tacgcaggcg ctcattctcg atcgacacgg cgacctcgtg ccgttcggca tcgccggcga    36900
gctctgcctc ggcgagacg ggctggcgcg ggaatacctg aatcagccgg cgctcactgc    36960
ggaacggttc gtgcccaacg cgttcagccc cgagccggcc gcccggctgt accgcacggg    37020
agacctcgcg cggatgcgcc gcgacggcaa catcgagttc cttggccgga tggatcgcca    37080
gatcaaactg cgcggcttcc ggatcgagcc gggcgagatc gagcatgccc ttcgtgcccc    37140
gggagacgtt ctggacgcgg tggtggagat ccgggtcgat gcccacatgg cccatctcgt    37200
cgcctacgtc gtgcgcgccg atgggagtca actcaccgga accgacctgc gcgaacgtct    37260
gaaggcgcgc ctgccggaat actgtgtgcc cgcgaagtac gtgttcctgg ccgaggtccc    37320
gcgaacggcg gccggcaagg tcgatcgctc ggcgctgcct gcatcggcga tcgaggcgcc    37380
cgcgcccgag ccgcgcgatc cgtcgttcct gaacgaggtc gaagcggagc tcgcgcgaat    37440
ctggagcgaa gtgcttcggc tgccctccgt gctcgcgacc gacaatttt tcgatcgtgg    37500
//
```

Figure 5b

```
CDS        complement(3..914)
           /blast_score=2e-66
           /blastp_match="AE004644.PA2177 Pseudomonas aeruginosa"
           /gene="regulator hybrid"
           /note="probable similarity to prokaryote sensory
           transduction proteins"
           /product="sensor/response regulator hybrid"
CDS        924..2168
           /note="none"
           /gene="ligase"
           /blastp_match="AP003013.MLR8297 Mesorhizobium loti"
           /blast_score=e-152
           /product="2-amino-3-ketobutyrate CoA ligase"
CDS        2207..3190
           /blast_score=e-151
           /blastp_match="AE008872.TDH  Salmonella typhimurium"
           /note="none"
           /gene="dehydrogenase"
           /product="threonine 3-dehydrogenase"
           /pfam_match="PF00107; adh_zinc; 1"
CDS        3373..4455
           /note="putative"
           /gene="methyltransferase"
           /blastp_match="AE001866.DR0026 Deinococcus radiodurans"
           /blast_score=1e-08
CDS        4546..4959
           /blast_score=2e-11
           /blastp_match="AP002997.MLL1617 Mesorhizobium loti"
           /gene="unknown"
           /note="pfam00263, GSPII_III, Bacterial type II and III
           secretion system protein, Expect = 7.8"
CDS        5176..6192
           /blast_score=5e-98
           /blastp_match="AF064070.PE20 Burkholderia pseudomallei"
           /gene="glucose epimerase"
           /note="putative "
           /product="UDP-glucose 4-epimerase"
           /pfam_match="PF01370; Epimerase; 1"
CDS        6331..14043
           /note="substrat AT, malonyl ; zinc depend dehydrogenase ;
           zinc dependent adenosine deaminase putative"
           /gene="PKS I"
           /blastp_match="AF285636.WCBR  2547 Burkholderia mallei"
           /blast_score=0.0
           /pfam_match="ketoacyl-synt"
           /pfam_match="ketoacyl-synt_C"
           /pfam_match="Acyl_transf"
           /pfam_match="SAM binding"
           /pfam_match="adh_zinc"
           /pfam_match="pp-binding"
CDS        14275..15408
           /blast_score=e-104
           /blastp_match="AF285636.WCBT Burkholderia mallei"
           /gene="acyl-CoA transferase WcbT"
```

```
                        /note="putative"
                        /pfam_match="PF00155; aminotran_1_2; 1"
CDS                     15436..16245
                        /blast_score=5e-11
                        /blastp_match="TTDEFFMT.FMT T.thermophilus"
                        /gene="formyltransferase"
                        /note="evidence experimental"
                        /product="methionyl-tRNA formyltransferase"
                        /pfam_match="PF00551; formyl_transf; 1"
                        /pfam_match="PF02911; formyl_transf_C; 1"
CDS                     16287..17384
                        /note="putative"
                        /gene="glycotransferase"
                        /blastp_match="AF285636.WCBD Burkholderia mallei"
                        /blast_score=e-99
CDS                     17427..18158
                        /blast_score=2e-82
                        /blastp_match="AF285636.WZM Burkholderia mallei"
                        /gene="ABC-2 transporter Wzm"
                        /note="putative"
                        /pfam_match="PF01061; ABC2_membrane; 1"
CDS                     18248..18847
                        /blast_score=7e-61
                        /blastp_match="AF285636.WZT Burkholderia mallei"
                        /gene="ABC-2 transporter Wzt"
                        /note="putative "
                        /pfam_match="PF00005; ABC_tran; 1"
CDS                     18952..20346
                        /note="putative"
                        /gene="glycosyltranferase"
                        /blast_score=e-101
                        /blastp_match="AF285636.WCBE Burkholderia mallei"
                        /pfam_match="Glycos_transf_1"
CDS                     20442..21167
                        /note="putative"
                        /gene="unknow"
                        /blast_score=1e-26
                        /blastp_match="AE009248.ATU3189 Agrobacterium tumefaciens"
CDS                     complement(21164..24301)
                        /note="tranporter domain"
                        /gene="unknow"
                        /blast_score=5e-29
                        /blastp_match="AE009122.ATU1658 Agrobacterium tumefaciens"
                        /prosite_match="PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1"
CDS                     complement(24351..27023)
                        /note="none"
                        /gene="unknow"
                        /blast_score=5e-29
                        /blastp_match="AP003581.ALR0267 Nostoc sp"
CDS                     complement(27806..29686)
                        /note="none"
                        /gene="cell surfarce protein"
                        /blast_score=2e-34
                        /blastp_match="AE010748.MA0851  2567 Methanosarcina
                        acetivorans"
                        /product="cell surface protein"
CDS                     complement(29535..30872)
```

```
                              /note="none"
                              /gene="cell surface protein"
                              /blast_score=2e-36
                              /blastp_match="AE010748.MA0851 Methanosarcina acetivorans"
        CDS                   complement(30848..32647)
                              /note="fragment"
                              /gene="O-antigen"
                              /blast_score=1e-72
                              /blastp_match="AF105060.RFBC Riftia pachyptila"
                              /product="O-antigen biosynthesis protein"
                              /pfam_match="PF00535; Glycos_transf_2"
        CDS                   complement(32574..35555)
                              /note="putative"
                              /gene="glycosyltransferase"
                              /blast_score=7e-46
                              /blastp_match="AE013462.MM2213 Methanosarcina mazei "
                              complement(35533..36598)
        CDS                   /blast_score=7e-37
                              /blastp_match="AE013462.MM2213 Methanosarcina mazei"
                              /gene="glycosyltransferase"
                              /note="putative"
        CDS                   complement(36516..37400)
                              /blast_score=8e-22
                              /blastp_match="AP003581.ALR0267 Nostoc sp"
                              /gene="ALR0267"
                              /note="putative ATP/GTP-binding protein, esterase fush
                              9e-14"
Sequence 37507 BP; 5531 A; 13507 C; 13011 G; 5458 T; 0 other;
```

```
gatcccacga tcggcgccag cggattgcgc agctcgtgcg cgagcatcgc gaggaactcg    60
gtcacgtggc ggccctcggc ctccagctcc tccatgcgct tgcggtgcga gaggtcgcgc   120
gtcaccttgg cgaagcccat gtgccttcca ctcgcgtcgc gaagcgccgt cacgacgacg   180
tcggcccaga agcgcgtgcc gtccttgcgc acgcgccagc ccgtgtcctc ggcgcggcgg   240
tgcagcagcg catcgtgcag ctcggcctcc gggcggccgg ccgcggcatc ctcgggcgga   300
tagaagagcg aaaagtgccg tccgatgatc tcttcggcgc tgtagccggt gatgcgctcg   360
gcgcccttgt tccagctggt cacgtggccc gcggggtcca gcatgtagat cgcgtagtcc   420
tgcacgccct cgatcaggag gcgcgagcgc tcctcgcttt gccgcagcgt gccgtcctcg   480
tcgcggcagcg cggtcaggtc gcgcgtgacc ttgccgaagc cgaccagcgt gccgtcctcg   540
tcgcgcagcg ccgtgatgat gacgttggcc cagaagcgcc ggccgtcctt ggcgggccag   600
cagccctcgt cctcgaagcg gccgtcctcg cgtgcgcgca gcagctcctt ggcgggccag   660
tcgcgcgcga tgtcctcctc ggtgtagaac gcggagaaat ggcggccgat gatctcttcg   720
gcgcgatagc ccttcaactg ttcggcgccc gcgttccacg aacgcacgta gccctcggga   780
tcgagcacga agatcgcgta gtccgtgacg gacccgacca gcagccggaa gatggcgtcc   840
tcgccggcca gcccgctcgg ggcgccgaag gcgaggttgc gcaccgaggg ccgcggaccg   900
gatgggcttg gcatggctat tagggtacaa cggtgcacgc aagttgcaat cggacaatca   960
gagataattg cttgatgaag atgccatttt tggaattctt acgcagggag ctggggacgc  1020
tcgaggccca gggcctgtac aagtccgagc gcgtgctcga ctctccgcag ggcgccagcg  1080
tgcaggtggc cgcccacgac gtgctcaact tctgcgccaa taactacctc ggcctggcca  1140
accaccggc cctggtcgcg gccgcgaagg agacgctcga gcgcgacggc tacggcatgg  1200
cgtcggttcg cttcatctgc gggacgcacg ccgtgcaccg cgagctggag acgcgcctcg  1260
cgcgcttcct cggcaccgag gacgcgatcc tgtacggctc ctgcttcgac gccaacggcg  1320
ggctcttcga gacgctgctg ggcgaggagg acgcggtgat ctcggacgcg ctgaaccacg  1380
cctcgatcat cgacggcgtg cgcctctcga cggcgacgc ggccggcgcg cgctacaaga  1440
acatggcgag cctcgaggcc cagctgaagg acggcatcgt cgcgaacctg ggcgccatct  1500
tgatcgccac cgacggcgtg ttctcgatgg acggcatcgt cgcgaacctg ggcgccatct  1560
gcgagctggc cgggcgctac cacgcgatgg tgatggtgga cgactcgcac gccgtgggct  1620
```

```
tcatgggcgc gcacgggcgg ggaacgcccg agcattgcgg tgtcgaaggg aaggtggaca    1680
tcctcaccgg cacgctcggc aaggccctgg gcggcgcttc cggcggctac acggcgggca    1740
agcgcgaggt ggtggcctgg cttcgcaacc gctcgcgccc ctatctcttc tccaacacgc    1800
tgatgcccgc catcgcgggc gcgtcgctca aggtgctcga tctcctcgag ggcggcggcg    1860
agctgcgcgc gaagctcgcg cgcaacgccc gccacttccg cggcgagatg acgcgcctcg    1920
gcttcacgct ggccggcgcc gaccatccga tcatcccggt gatgctgggc gaggccccgc    1980
tcgcgaagga gatggcggac cggctgctga aggaaggcat ctacgtggtc ggcttctcgt    2040
ttcccgtggt gccccgaggg caggcgcgca tccgcacgca gatgtcggcc gcccatgaac    2100
cgaagcacgt cgatcgcgcc atcgccgcct ttgccaaagt cgggcgcgat ttgggagtca    2160
ttgcttgaga acccttcga agacgaagcg cgagcccggc atctggatgg tcgagtcgcc    2220
caagcccgtg gtcggccaca acgacgtgct gatccgcgtg aagaagaccg ccatctgcgg    2280
caccgacatg cacatcttca actgggacga ctggtcgcag aagaccatcc cggtgccgat    2340
gacggtcggc cacgagtacg taggcatggt ggaggccatg ggccaggagg tgcgcggcct    2400
gcaggtcggc cagcgcgtct ccggcgaggg ccacatcgtc tgcggccatt gccgcaactg    2460
ccgcgccggg cgccgccacc tgtgccgcaa cacgcagggc gtgggcgtga accgccccgg    2520
cgcgttcgcg gactacctgg tgattcccgc ggagaacgcg ttcccgattc ccgacgacat    2580
ccccgacgag atcgcctcga tcctcgatcc gttcggcaac gcggcgcaca ccgcgctctc    2640
gttcgacctg gtgggcgagg acgtgctgat caccggcgcc gggccgatcg gcatcatggc    2700
cgcggccatc gcgcgccacg tcggcgcgcg ccacgtcgtg atcacggaca tgaacgacta    2760
ccgcctggcg ctcgccacga agatgggtgc cagccgcgcg gtgaacgtat cgaaggaaaa    2820
cctgaaggac gtgatgcgcg agctcggcat ggtcgagggc ttcgacgtcg gcatggagat    2880
gtcgggcgtg ccctcggcct tccgccagat gctcgacacc atgaaccacg gcggcaagat    2940
cgccatgctc ggcatcccgc ccagcgaggc cgcgatcgac tggacccagg tgatcttcaa    3000
gggcctggtc atcaagggcg tctacgggcg cgagatgttc gagacctggt acaagatgat    3060
cgccatgctg cagagtggcc tcgacctctc gcccatggtc acgcaccgct tcgacgtgcg    3120
cgagtacctg aagggcttcg agacgatggg ctccgggaaa tccgggaaag tcgtcctgtc    3180
gtgggattag cgagggccac ggatcgcggc ccggttcccg cggcggcgat cgtgcaggcg    3240
ggttggcagc acgccccggg atttccgccg tggctcacgt gcgacttcgc cgtcgacgca    3300
ccgatcgcga tcgacgtcga cgtgagcgtg tcgctgctcg atgcccaggc cacctcgctg    3360
tggacgcagt cgatggccgg cagcgggcgc gacaccccgt ggctcacgat cggcacgtat    3420
cgcgcctccc tgcgcctgga gccgttccgc ttcccggccg cggcggcctc ggtcgagttc    3480
gcgctctcga tgcgggtgca aggtgagcgc cgcgtcgtgg ccaccgcagc gggcgtggtg    3540
cccgcggggcg cacccacggg cgtggcgcgc gcggcctggc acctcgaggc cctggatggc    3600
acgccggcgc tcgagacgct ggcctggtcg gatcccacgc acagctggtt ctcccagcat    3660
ttcgaccatg cgtcgcgcac gtgcgtcgag tacctgggcg ccggatggcc cgggtggcgc    3720
gggcgcgtgc tggatgcggg atgcggcgac ggcatcacgg cgctcggcat cgcgctgcgc    3780
tatgcgcccg agcaggtggt gggcatcgat ccggggcgct gctaccgggt gcttcccgac    3840
atcctcgagc gccacgccct cggccacctc gcgctgccgg acaacctgca gttcctcccc    3900
gcggatgcca acgcgctccc gttcgccgac agcagcttcg acgtcatcgt ctcgtggtcc    3960
gcggtcgagc atttcgtggg cggctacctg ccctcgctcg ccgaggcgcg gcgggtgctc    4020
aagcccggcg gcctgctggt catccatccc gagctctact acaccgcgca ccacggccat    4080
cacctcggcg agtacagccg cgagcccttc ttccacctgg tgaagtcgcc cgacgaggtg    4140
cgcgacatcg tgttcggcgc ggacgtgagc ctccacgacc gcggcggcat cgcgccgacg    4200
cgcgcggagc actggcgctg gtacaacgag ctcaatcgca tcacggccgc gcagttcgag    4260
gacgagttgc gcgcgctcga cttcgagccg tggcgcctgg cactgcgcgc cgagcgctg    4320
gtggagtaca gccggagct gctgcgctac cgcttcacgg agctgggcgt ctccgagctc    4380
tacgtggcct gcatcaaccg caagcccgtg tcagatggtt cctacaggac gcatgggcag    4440
gcggcattcc ggtaaggtcc gccgggacgg gaattgaagg cgcaagcgcg ttgttatccc    4500
cgtcactccc ggactgccgg gatcaccctc acaggagaac ccccgatgaa gacccagacc    4560
caggccgtgc tcgccgcggc cctgctgtcc cttgccggca ccacgctggc cgcgccgaa    4620
gccgtcacgc tgacgggcgc gagcgaagtc ccgcccgtca cctcgagcgc gaagggcagc    4680
ggcacggtgg tcgtgaaggc cgattgcacg gtcaccgcga agatcaccgt caccggcatg    4740
accgccaccg cggcccacat ccacgaaggc aaggccggcg ccaacgcggg cgtggcagtc    4800
ccgttcgtga agaccgcgga cacacggttc gaggcagccc cgggcgccaa gatgaccgag    4860
gcgcagtgcg ccgcgtacaa ggcggcggc acctacgtga acgtgcacag cgaggcccac    4920
aagggcggcg aggtccgcgc ccagctggcc ggcaagtagg cagccgctg caaaaagaaa    4980
aagcccgcgc aagcgggctt tttcgttgaa ggctttggtc ggggcgaaag gattcgaacc    5040
```

25/39

```
ttcgacccc  tgcaccccat  gcaggtgcgc  taccaggctg  cgctacgccc  cgaccgaact    5100
tgaaattata  cctgcatcgg  ggtcgcagcc  gccatgaatc  gaaacgcccc  gggcttcggg    5160
tagcatccgc  ccatggcgac  gaccctcgtg  acgggtggtg  caggctacat  cggcacccac    5220
atcctctgcg  cactggccca  ggccgggcga  cgcagcatct  gcatcgacaa  ctactccaac    5280
agctcgccgc  gttccatcga  acgcgtgctg  cagatcgcac  ccggttgcgt  cgaggccttc    5340
gacgtggaca  tccgcgacgc  cgacggcatc  cgcaaggtga  tcgcagggcg  cgacgtggac    5400
agcgtgatcc  acctcgcggg  actgaaggcg  gtcggcgaat  cggtcgagca  gcccgagcgc    5460
taccacgaca  acaacgtccg  cggcaccgag  agcctgctcg  cggcgctcgc  cgactcgccg    5520
gtgcgcaagt  tcgtcttcag  ctcctcggcc  accgtgtacg  gcctcgccga  gaagatgccg    5580
atcgacgagg  acgcgcccac  ctcgccgcag  agcccgtacg  gccagaacaa  gctcgacatc    5640
gagcacatgc  tggtcgccct  cgccaagcgc  gacccctcct  ggcgcgtagt  gaacctgcgc    5700
tacttcaacc  cggtcggcgc  acacgagagc  gcgctgatcg  gcgaggaccc  ggccggggtt    5760
cccaacaacc  tcatgcccta  cgtgtgccag  gtcgcctccg  ggcgactgaa  ggagctgagc    5820
gtctacggca  gcgactaccc  cacccggac  ggcaccggca  tgcgcgattt  catccacgtc    5880
tgcgacctgg  ccgaaggcca  cgtggccgcg  ctggaagcac  tcgagcgcgc  ggcccccggg    5940
accgtcctca  cggtgaatct  cgggacgggg  ctcggctaca  ccgtgctgca  attgctcgaa    6000
accttcgagc  gcgtgaacaa  actaagcatt  gcccggcgca  tcgtgggggcg  gcggccgggc   6060
gacgtcgcga  tctgctacgc  gaacgcgggc  cgcgcgaagg  ccgtgctggg  ttggaccgcg    6120
aagcgcggaa  tcgaggagat  gtgccgcgat  gcctggcgtt  ggcaggaaag  gaatgcaaag    6180
ggctacggtt  gacggattgc  aaatcatggc  cttggcgacc  cttgctgcaa  tgcggcatga    6240
aggggcggt  tcttgtaaga  tccaaggttc  gggatgcgtc  caggacgtca  aaaacagcac    6300
ccggcgttgg  aaggaagaag  gcaaccctag  tgttgtcatg  gctccagacg  ggattgcgcg    6360
ccgagtaaga  atcgtcagtc  gagaccaact  cgcctccggg  tgcaacgaat  cgtgagccga    6420
agcaacacct  cgcatcgcgt  cgccgtcgtc  ggcattgcgt  tccgccttcc  cggcgctacg    6480
cgcgaatccc  tctggccctc  gctcctcgag  gggcgcaacc  tcgtcaccac  cgtcgatccc    6540
gcgcgctggg  cgaaggagcc  gtactaccac  ccgcgcaagt  ccgagcccgg  caccagctac    6600
tcgttcgccg  ccggcaccat  cggcgacatc  agcggcttcg  acgcgtcgtt  cttcggcatc    6660
tcccgcgcg  aggccgcgca  gatggatccg  cagcagcgcc  tgctgctcga  gctctcctgg    6720
gaagccctcg  aggccgcggc  ggtgaagccc  tcgaccctgc  gcggcagcga  ttgcggcgtg    6780
ttcgtgggcc  tgtccaccgt  cgactacagc  tacggcctcg  cctacgacct  ggcgtccatg    6840
gacgcgtcgt  cggccacccgg  caacaccgcg  agcatcgcgg  ccaaccgcct  ctcgtatttc    6900
tacgacctcc  acggcgccag  catggtcgtg  gatacggcgt  gctcgtcctc  gctggtggcg    6960
ttccaccagg  cctgccagtc  gatcgccatc  ggcgagacgc  gccaggcgct  ggtgggcggc    7020
atcagcctgc  acctgcatcc  gtacgggttc  atcgcgttct  ccaaggcgtc  gatgctctcg    7080
aagcgcggcg  cctgccgcgt  gttcgacgcc  agcggcgatg  gctacgtgcg  ctcggaaggc    7140
gcgggcgtgc  tggtgctgaa  ggacctcgac  cgcgccgtgg  ccgacggcaa  cccgatcctc    7200
gccgtcgtcg  ccgcgagcgg  cgtgaatacc  gcgggccgct  cctccagcct  caccgtgccc    7260
agcgtcgatg  cccaggctgc  attgctgtcg  gatgtgtacg  agcgcgccgg  catcgacccg    7320
tccgagatcg  actacctcga  ggcccacggc  accggcaccg  tggtgggcga  tccgatcgag    7380
acccgcgccc  tcggcatcgc  gctcgggcgc  ctgcgcccgc  gccaccagcc  gctgccgatc    7440
ggttcggtga  agagcaacgt  cggccacctc  gaggcggcct  cgggcatggc  gggcctggtg    7500
aaggcgctgc  attgcctcaa  gcaccgcgcg  gtgccgccca  ccatccacct  cgattcgccc    7560
aaccccaaca  tccacttcga  cgagtggaac  ctgaagccgg  tcctcgagcc  gctcacgctc    7620
catccgtaca  agaagctcgt  cgtcggcgtg  aactcgttcg  gcttcggcgg  cgccaacgcg    7680
cacgtgatcc  tcgagagcac  gcccacttcg  cccaaggcgg  tcgaggcccc  ggccggcgag    7740
gtggcgctca  tgctgagcgg  ccacgacgac  gcggcgctgc  gcgacgtggc  ggcctcgtac    7800
gcgaacttcc  tgcaggcgag  cgagcaaccc  gcgctgcacg  acatcgcgta  caccaccgtg    7860
cactccgcgc  actggcacca  gctccgcgcc  atggtggtcg  gcaccgaccg  cgccagcatc    7920
gccggcgccc  tcgcgcaatt  cgccaagggc  cacaacacgc  cggcgtcgt  gagcggccgt    7980
gcgctgccca  acgcgtccgg  ccccgcattc  gtgtattcgg  gcaacggcgc  gcaatggccg    8040
gcaatggggc  gcgccctgct  cgccggcgag  ccggtgttcc  gtgccgccgt  gcaatcggtc    8100
gacgcgctgt  tccgcccgct  cggcggctac  tccatcctcg  agatgctgga  gcaggaggtc    8160
cccgccgacc  agctggaccg  caccgagatc  gcccagccgc  tgctcttcgc  cgtgcaggtg    8220
ggcttgaccg  agctgctgcg  ccactggggc  atccagccca  cgcgcgtcac  cggccacagc    8280
gtgggcgagg  tcgccgccgc  ttgggcgagc  ggcgcgctga  cgctcgacga  tgcggcgcgc    8340
gtgatctacc  accgcagccg  ctaccagggc  gagaccaagg  gcgcgggcgc  gatgaccgcc    8400
gtcggcctcg  ccgaggacgc  cgcgcgccgc  cagatcgagg  cgctcgggtt  gaagggcgag    8460
```

```
gtgaacgtcg cgtgcatcaa cagcacgcgc aacgtgacgc tcgccggcac gcgcgccggc    8520
atcgaggcgc tcgaggccga gctcacgcag cgcaaggtgt tccaccgccg cctcgacctc    8580
gactacgcgt tccacagccc ggcgatggac ccggtgcgcg acgggctcgt gagcgcgctg    8640
cgcggcctca cgccgcgcgc cacccgggtc gcgttccatt cggcggtcac cggctcgccc    8700
gccgcgggca accagctcga cgccacctac tggtggagca acatccgcga gccggtgcgc    8760
ttccaggccg cgatccgcgg catcgccgag tcgggcgtca atgtcttcat cgagatcggc    8820
ccgcacccga tcctcaagaa ctacatcaac gacggcctgc gcgccgcctc gatcgagggc    8880
cgcgcgctgg tcacgctgca gcgctcggcg agcgatcgcg cggcgatccg cgccgccgcc    8940
caggaagtcg tgatcaccgg ctgcccggtg gagaccgcga agatgctccc cgcgcagggc    9000
cacttcgtcg agctgccgcc ctacccgtgg cagcgcgagc gccactggcg cgcgcccacc    9060
tcgcaggcct acgacctcat ccagcacggc aagcagcacc cgctgctggg ctatcgcctg    9120
cacgagaacg acttccagtg ggaaaaccac atcgacaccg cgctctatcc ggcgtacgcg    9180
gaccatgtcg tcggcggcgc cgtggtgttc cccgcggccg gcttcgtcga gatggcgctc    9240
gccgcctcgg cgatcgcgct gggcggcgac gcgcacgaga tcgagacgct cgagatccgc    9300
agcgcgctgc tgctcgagga ctccacgtcg aagaccgtgc gcttcgcgct ggaacccgat    9360
ggccgcttca caatccgcag ccgcgctcgc ctgagcgagg atccgtggca actccacgtg    9420
gtcggcaagc tggtcggcac gcccaccgag cttccgcgcg cgccgcgcca gtcggcgccg    9480
tcgcgcaagc ccgatgtgat cgcggcccag cactacgaag gcgccgcgaa ggccggcctc    9540
gcctacggcc cggccttcca gtcggtctcc aaggtgtggc tcgacgccgc ggacccgggc    9600
agcgcctacg cgcgcctgat cctccccaag ccgattcgcg gcgagctggg cgtgatgcac    9660
ctgcaccccg cgtcgctgga cggctgcttc cagctgctgg tcgacctgct gcgcgccgag    9720
gcctcccggc acgcgcagat cgccttcgtg ccgatcctgg tcgggcgcac gcgcctctac    9780
ggccctgccg ccctcgtcac cgcctgccgc gtccgctga ccgcgcgcag cccgcgctcg    9840
ctggtcgcgg acttcgagct ctacggcccc aacggcgacg tggtggccgc gctcaccggc    9900
gtgcgcttcc gcggcgtcca gctgaagcag ccgcagggcg cgcgcctgcg ctacctgtcg    9960
taccgcgcca tcccgcgcct gaatggcggc gacacgccgg gcgccaagcc gctgcccctc    10020
gcgttgctgg ccgcggcgtg cggcgagcgc tggcattcgc tctcggccgg ctcgcgcaag    10080
cgctactacg atgaggtcga gccgctctcc gacgtgctgt gcagctcctt cgccgagcac    10140
gcgctgcgaa agcggatcgg cgaggccccg ctcatcgacc cggccgcgct cgccgcgtcg    10200
acgccggccg agcattggcc gctcctcaag cgcgcgatcc agatgctggt cgaggaccag    10260
ttgctcgagc ccgccgaagg cggctggcgc tggagcccgg tcaccgaatt gcccgacgcg    10320
caggagagct ggatcacgct gctgcgcgac tacccggacg aggccggcca gctgatcgcg    10380
ctcgggcgcg ccggcgctca cctcgcggaa gtgttcgccg gcaaggaagc gtccttgccg    10440
agcctgttgc cggccacgca ggcgaacgtg ttcgcgcgcc tgtgctcggg cgcacccgcg    10500
ttcgccaacg ccgggcgcgc catcgccgac acgctggcgc tcgccctcga gcgcctgccc    10560
gcgcaccggc cgctgcgcgt gctcgaggtg acgccgtgcc gctccgagat cacgctgcgc    10620
gtcctcgagg ccgtggacct cgaccgctgc gagtaccacg tcggctgcac cacggacgaa    10680
gcactgggcg agtacgaagg cgtcctcgat gccattccgc acgtcgagac ttgcgtcgtg    10740
gacctgaaga tcccgggcat gggcctgaag gcgccgaacc agggcccgtt cgatgtcgtg    10800
atcgtttccg atggcctgct cggtgcggcc gatccggatg cggccgcggc gcacctcgcc    10860
ggcgcgctcg cggaagacgg gctgctggtc atggtcgcgc agcatccgtc gcgctggacc    10920
gacctcctct tcggcgcgga cccggcgtgg tggacgctcg ggccctcggg ctcgcagcgc    10980
tcgcgcctgc gttcgccgga cgagtggcgc atcgcgctgg cccacctcgg cttcgggccc    11040
accgtcgcgg tgcccgagct gcccggcctc gagcgcgggct cgtacgtcct catggcgcgg    11100
cgcgacggcc cggtgccgca ggctgccgag cgcagcgttc ccgccaagtc gggctggatc    11160
ttcgtgcagg aagcgggcgg ttattcggcg gcgttcggcg cctgcgtgca ggcggatctg    11220
cgcgcacgcg ccacgcggt gttcgagatg gcgcccgatt ccgtcgaggg ggccgagatg    11280
gcgcagcttg cgcgcgcgcg ccagacgctg ggcgacatca gcggcatcgc cttcctggcc    11340
ggcctgcccg acgacgcggc cgccacgggc gacgcggccg agcgcctcgc gcgccagacc    11400
gcgcgctgcg ccgcgctcgc gcgcctgctg cgcgaatgca aggcgagtgc catgacaccc    11460
gcggtgtgga tcgtgacctc gggcgcggcc tccagcctgg cgcgcaacgt ccccgccctt    11520
gcctcgcgcc gccgccctcc cgatccggac gaggcgatcg tatggggggtt cgcgcgcacc    11580
gcgatgaacg aattccccga gctgcgcctg cgcctcgtgg actgcccgaa cccggagaac    11640
ctcgagcgca acgccgcctc gctggtgcag gagctgctct cgcccaccgc cgacgacgag    11700
gtggtcctca ccggcgccgg ccgctacgcg ctgcgcctgg gcgcggtgcc gccgcccgca    11760
tccgtgcgca ccgctaccca gggcatcacg cgcctcgact tcgccgcgcc cggcccgctg    11820
aagaacctcg cgtggcgcgg cgacgtgcgc cgcaagccgc gcgcgaacga ggtggaaatc    11880
```

```
gaggtgcgcg ccgcgggcct caacttccgc gacgtgatgt acgcgatggg cctgctctcc   11940
gacgaggcgg tcgagggcgg cttcgccggc gcctcgctcg gcatggagct ctcgggcgtc   12000
gtcaccgccg tcggcaagga cgtgacctcg gtcgcgcgcg gcgacgaggt gctcgcgttc   12060
gcgccctcgg cgttcgccac gcacgtggtc accaccgccg attcggtcgc gaagaagccc   12120
gccggctgga cgttcgagtc cgcggccacc atcccgaccg cgttcttcac cgtgtactac   12180
tcgttgaagc acctggccca gctgcgcgag ggcgagcgcg tcctgatcca cggcgccgcg   12240
ggcggcgtgg gcatcgccgc gatccaggtc gccaagtggc tgggcgcgga gatcttcgcc   12300
accgccggct ccgacgagaa gcgcgacttc gtgcgcctgc tgggcgcgga ccacgtgctc   12360
gactcgcgca cgctcacgtt cgccgacgac gtgctgcgca tcaccaacgg cgagggcgtc   12420
gacgtggtgc tgaactcgct ctccggcgag gccattgcgc gcaacctgcg cgcgctgcgc   12480
ccgttcggcc gcttcatcga gctcggcaag cgcgactact acgagaacac gcacatcggc   12540
ctgcgcccgt tccgcaacaa catctcgtac ttcggcgtcg atgccgacca gctcatgaag   12600
gagcggcccg acctcgcgcg ccagctcttc aacgagctga tgcagctctt cgagcagggc   12660
gtgctctcgc cgctgcccta ccgcgccttc gccgccaccg aggcggtgga agccttccgc   12720
tacatgcagc actcgcgcca gatcggcaag gtggtgctgt cgttcgccga cggcgtgaag   12780
gcgcagcccg cccggcccgc cgtgccgcgc gagctggcgc tgtcgccgaa cgccacgtac   12840
ctcgtcaccg gcggcctgtc gggcttcggc ctgcgcaccg cgcagtggct ggtggacaag   12900
ggtgcccgcc acctggtgct ggtctccaag agcggcgccg agtcgctcga gaccaaggcc   12960
gcggtggccg acctcgaggc ccagggcgcc acggtgatcg cggaggcgtg cgacatcacc   13020
gaccgcgcct ccgtgcaacg gctgctcgcc gaggtcgcgg ccgcgctgcc gccgctgcgc   13080
ggcgtgatcc acgcggcggc cgtgatccag gacggcttca tcaccaacat gacggccgac   13140
cagatccgcg acgtgctggc gcccaaggtc ctcggcgccc gccacctgga cgagctcacg   13200
cgcggcgcca agctcgattt cttcgtgctg tactcgtcgg cgacgacgct gttcggcaac   13260
ccgggccagg ccaactacat cgcggccaac tgcttcctcg aggcactcgc caaggcgcgc   13320
cgcgtcgagg gccagccggc gctgtgcgtg ggctggggcg cgatcggggga cgtgggctac   13380
ctcgcgcgcc acgagaaggt gaaggaagcc ctgcagacgc acatgggcgg caccggcactc   13440
gagtcggagg ccgcgctcgc ggtgctggag cagctcctgc tcgccgacgc ctcgggcctc   13500
ggcgtgctcg acttcagctg gcgcaacctc cgccgcttcc tccccaccgc cgcctcgccg   13560
cgcttccgcg agctgggcgc gcgcggcgac gacgcgaagg acgacgacga ccggcgcggc   13620
gagctgcgcc gcctcgccaa cgagctggac gcctccgagc tctcggcgat gttcaccgac   13680
ctgctgcgcc gcgaggtggg cgagatcctg cgcatcccgc cggaccgcct cgatacgcgc   13740
cagcccttgc aggaaatggc catggattcg ctgatgggcg tggagctgct gaccgcggtg   13800
gaggcgcgct cgggcgtgaa cctgccggtg atgtcgctct ccgagcagcc gtcgatcgag   13860
aagctggtcg accgcatcgt gcgcgcgctg aaggacccga acgccggcac cgaggccgaa   13920
tcgcgcagcg accagatcga gcgcgtcgcc gcgcagtacg cgcccgagct cgactcgcgc   13980
caggtcgagg agctcacgca ggcggtcgat gacgcgcaag ccacgcagca gggacgtccg   14040
tgaaggcgcg caacctccgc ggcctctcgc aggcagtgaa ggagcagctg atccagcgcg   14100
tgctcgagca gcgcgtgcgc cgggtcgaga aggacaagcc ctcggcggtc gagtccgcca   14160
tcaacgcctt ccgccccgac cccgatgccc tgggccacat ctcggagcgc tactgccgct   14220
tcgacatgca cccggtctac cagcagatgc agctggtgaa ggaaggtgcc gcgaagctgg   14280
gcattcccga tccgtacttc cgcgcgcacg acggcgtggc cgcggatacc acgcgcattg   14340
ccgggcgcga gtacatcaac ttctccagct acaactacct cggcttctgc ggcgacccgg   14400
ccgtggatgc cgcggccaag caggccatcg acgaatacgg gaccctcggt ctcggccagc   14460
aggctcgtgt ccggcgagcg gccgctgcac cgcgagctgg agcgcgccat cgccgaggtc   14520
tatggcgtgc acgacgccgt ggccttcatc agcggccacg cgaccaacgt ctccaccatc   14580
ggccacctgt tcgggccgcg cgacctcatc gtgcacgacg cgttcgtgca caacagcatc   14640
ctgatgggca tccagctctc gggcgccaag cgcatggcct ccccgcacaa cgactggcgg   14700
gccctggacg agctgctgcg cgcgcagcgc cgccacttcg agcgcgtgct gatcgtgatc   14760
gaggggcgtct acagcatgga cggggactac cccgagctgc ccgaattctc gcgcctgcgc   14820
cgccgccacc gcgcgttcct gatggtggat gaagcgcact ccttcggcgt gatgggcccg   14880
cgcggcttcg gcatccgcga ccacttcggc atggccggcg acgaggccga catctggatg   14940
ggcacgctct ccaagacgct cgcctcgtgc ggcggcttca tcgccggcga gaaggcgctg   15000
gtcgagcacc tgaagttcgc cgcccccggc ttcctctaca gcgtcggcat ggccccgccc   15060
gtggcggccg cggcgctcgc cgcgctgaac cgcatgcgcg aggagccggc cgcgcgtggcc   15120
gccctgcagg cgcgcggacg cttcttcctc gaggccgcgc gcgccgcggg catcgacgtg   15180
ggcctctcgc agggcatggc cgtggtgccc gccatcaccg gcagctcgat ccgcgccggg   15240
cgcctggccg aggcgatgtt ccagcgcggc atcaacgtcc agccgatcgt ctacccggcc   15300
```

```
gtgcaggaga actcggcgcg cctgcgcttc ttcgtgagcg ccacgcacac cgaggagcag    15360
ctgcgcttca cggtccgcga gctcgccgac gcctggcgca agctctgagt ggcgggcccc    15420
cggctgcgcg tcgtcgtctg cacccacggc ggcctgaacg gcgcgctggt gctctcgcgg    15480
ctgctggccg cccccacgct cgaagtctcc gcgctggtga tctccagccg cgcgcgcggt    15540
gcccatgagt cgtatggccg tgccgcgctc ggatacgtgc gcgcgagcgg cgtcgcctat    15600
gcgctgtacc tctggtgcgc cacggcgctc gccgacctgc tgctgcgcgg cacgtccgag    15660
ggcccggtcg cgcgcatcgc cctcgcgcgc ggcatcccgc tcctcgccac gccgcgcgtg    15720
aacgatgcca ccgcgcgcgc cttcatcgcg ggggcggcgc cggacctgat cgtctccgcg    15780
ttcttcaacc agcacatcga cgccgacgtg gctgcgctcg cgcgcgtggc cgcggtcaac    15840
atccaccgt cgccgctgcc gcacttccgc ggcgtggatc cggtgtcctt cgcgcgcctg    15900
cgcggtgccg agcgccacgg cgtgagcgtg catcgcatcg aacccggctt cgataccggc    15960
gcgctgctcg cccaggaaac cgacgtcgag gcccccggca gcgtctttgc cgccaccgcg    16020
gcgctctatg accgcggggc ggcgctcctc gccggccgtg ccgcggcgct ggccgccgac    16080
ccgcgcggaa cccccagcc ggcggggggc tcctacgact cctggcccac ccgcgcccag    16140
gtcgccgcct tccggcgcgc cggcgggcgc ctgctccgcg cccgcgacct gtggcgcctc    16200
gcgcgccggg ggccggccgc tttcgtaata gaatcagcgc ggtagcggcg actcacgccg    16260
cgcttttccc cgaccccccg aggcccatga aacactggct gaagcaacac gcaatcttcg    16320
cgctcaccgt cctgctgccg accgtggccg cgatcctcta cttcggcctg atcgcctccg    16380
acgtctacat ctccgaatca cgcttcgtgg tgaggagccc ccagcgccag gtgcagaccg    16440
gcctggtggg cgccctgctc tcgggcaccg gcttctcgcg ctcccaggac gacacctact    16500
cggtgcacga cttcatcacc tcgcgcgacg cgctgggcga gctggacaag aagctcgccg    16560
tccgcaagct ctacacggcc gccaacatcg acttcatcaa ccgcttcccg gggctcgact    16620
gggacgacag cttcgaggcg ttccaccgct actaccagaa gaaggtcacg atcgacttcg    16680
acaccgcgtc ctcgatcacg gtgctgcgcg tgcgcgcctt cgagaaggcc gactcgcggc    16740
gcatcaacga cctgctgctg cagatgggcg agcgcctggt gaacgagctg aacgagcgca    16800
gccgccagga cctgatccgc ttcgcgcagg ccgaggtctc gctcgccgag gacaaggtga    16860
aggatgccgc gctggcgctc tccgccttcc gcagcaacca gtcggtgttc gagcccgacc    16920
gccaggcctc gatccagctg cagggcgtgg ccaagctcca ggaggagctc atcgccaccg    16980
aggggcagct ggcacagctg cgcaagctct cgcccgacaa cccgcagatc ggcgcgctcg    17040
agaacaagtc ggcggcgctg cgcgtggcga tggcgcgcga gtccgccaag gtgaccggcg    17100
gcagcggctc gttcagcgcg cgcgccccgg cgttcgagcg cctcaccctc gagaagggct    17160
tcgccgaccg ccagctgggc gttgccctca ccgcgctgga gaccgcgcgc agcgaggcgc    17220
agaggaagca gctctacctc gagcgcatcg tgcagcccaa cctgcccgac gaatcggtcg    17280
agccgcgccg catccgctcg atcttcaccg tgttcgtgct gggcctggtg gcctggggcg    17340
tggtgagcct gctggtggcg agcgtgcgcg agcacgtcga ctaggccgtg gagcagaacc    17400
cttccctcct gcgcgccctc ggggtgcagc ggcgcgtgct ctacgcgctg ctgatgcgcg    17460
aggtcatcac gcgcttcggg cgcgacgacc tcggcgtgct gtggctggtg gtcgagccga    17520
tgatcttcac gctcggcgtg acggccctgt ggaccgcggc cggcatgaac cacggctcgt    17580
cgctgccgat cgtggccttc gcggtcaccg gatactccgc ggtgctggta tggcgcaact    17640
gcgcctcgcg cgcctcgatg gccatcgagg ccaacaaggg cctgctcttc caccgcaacg    17700
tgcgcgtgat cgacgtgttc gtgacgcgca tcatcctcga gatcgcgggc gccacggcgt    17760
cgttcaccgt gctgggcatc ttcttcatct ggatcggctg gatgccgatg ccgatcgaca    17820
tgctcaaggt cgcgttcggc tggttcatgc tgtcgtggtt cggcgccagc ctggcgctcg    17880
ccatcggcgc cggcaccgcg tactcgctgg tggtcgagcg gctgtggccg ccgacggcct    17940
acctgctgtt cccgctctcg ggcgccgcct tcatggtcga gtggctgccg gagaaattcc    18000
agcagttcgt gctcttgctg ccgatggtgc acggcaccga gctgctgcgc gacggctact    18060
tcggcaacgc ggtgcgcacg cactacgacg tgggctacat ggcaacggtc tcgctgggcc    18120
tgtcgctgct cggcatgcac ctggtgcgca tggcggccaa gcgcgtggag gccgcgtgat    18180
ccgggtcgag aacgtgacca gatgtactca cgcgcctcgg cccagaacgt gctgacgcgt    18240
cagcttcgag ctcggccgcg ggcgcaacat cggcatcctc ggcaagaacg gcgccggcaa    18300
gtccacgctg atccgcctga tcagcggcgc cgagagcccc accaccgggc gcatcgtgcg    18360
cgagatgagc gtgtcctggc cgctcgcctt cggcggcgcg ttccagaccc acctcaccgg    18420
gctggacaac ctgaagttcg tctgccgcat ctacggcgtc gactaccgcg acaaggtcgc    18480
cttcgtcgag gacttcaccg agctcggcgt gtatttccgc gagccggtgt tccactactc    18540
gcacggcatg atcacgcgcc tcgccttcgc gctgtcgatg gcggtggagt cgactgctt    18600
cctgatcgac gaggcgatgg tggtgggcga cgcgcgcttc cacgagcgct gccacgtcga    18660
gctcttccac aagcgcaagg accgcgcctt catcctcgtc acccacgacg ccaaggtgat    18720
```

```
caagctctat tgcgagagcg cctgcgtgct gcacgagggg cggctgctgc ccttccccac   18780
cgtggacgcc gcctacgagt tctacatgaa cgaggtcatg caggacctcg ccccggaggt   18840
tgcctgaacg ccccccgcgcg gccgctgcgc gtgctcctcg agctgcggcc ggcgttcgac   18900
ggccacgccg gcatcccgca ggaggcacgc ctgctgttcc gcggcctgcg catgatcgag   18960
ggcatcgacg tggagggcct gctgcagcac agcggccacg tgctcgccaa gggcctgccg   19020
ccgcgcggcg gcggcgacct cgcaccggac cgccagctga accgcctgtc gcgcgtcgtg   19080
gtctccacca agcaggagct caccaacgcg cacgtggcga cggcgctgat ggccattcgc   19140
cagctgctgg gcgcgcgcga gaccctctac cgcttcgacg cggcgcactt ccgcgacttc   19200
gtctggcagg cgctcttcgc gcgcacgctg cacgccgcgg acttcgactc ggtcacccgc   19260
gccggctttc gcgtcgcgcg cgtgccgtgg acgggcatgc accgctgcgc gctggtcacg   19320
cgcaagctcg gctacacgct gcatccgcgc atcgacactt ccgatttcga cgtgctgatc   19380
gccgagacgc cctacccggc gcgggtcgcg gacggcacgc gcctcgtggt gcgctaccac   19440
gacgcgatcc cgctcctcat gccgcacacc atctccgaca tgtcctacca ccaggcctcg   19500
cactaccggg cgctgcgccg caacgtggcc tcgggcgcgc acttcgcgtg cgtctccgaa   19560
gccacgcgca aggacctgct ctcggtcttt cccgaggtcg aggcgcgctc cagcaccatc   19620
cacaacatgg tctcgcacca ctacttccgc gagaccgccg acgccgggcg catcgaggag   19680
atcctgcgaa cgcgcgcgag cgagcgcatc aagggcgcgc ccaacggagc cgccaccgcc   19740
gtcgcacgat cgccctcctt cctcgcgcgc gccgcgcaac ccagcgaccc cggctacctg   19800
ctcgccgtgg cgacgatcga gccgcgcaag aaccacgcgg gactgatcgc cgcgtgggag   19860
cacctgcgca cctcgcgctt cccgggcctg cgcctggtgg tggtcggcat gcccggctgg   19920
ggctacgagg cgctggtagc caagttcaag ccgtggctcg cgcgcggcca gctcttcgtg   19980
ctcgaggacg tgcccgcccc cgagctgcgc ctgctctact cgcacgcgcg tgccacggtg   20040
tgcccgagct tcggcgaggg cttcgatttc tccggcgtgg aggcgatgcg ctgcggcagc   20100
ccggtgatcg cctcggagat cgccgcgcac cgcgaggtgt accgcgacgc cgccgagtac   20160
tgcagcccct attccgtggc cgacctcgcc gaggcgatcg ggcgcgtcat cgacccggcg   20220
gcgacgggcc tgcgccaggc gctggtcacg cgcggcaccg aggtctcgca gcggtacacg   20280
cccgaagcca tcctgccgca gtggcgcgag tacctgctcg gcacggtgcg cgcggaggcc   20340
ccgtgagcga tccgaaggcc accgaccagg cgtgggagga gtggggccag cgcgacccct   20400
acttcggcgt gatcaccaac ccgcgcttcc gccgcgggca gatggacgag gaagcgaagc   20460
gcgaattcct cgcgtccggc cgcgtccatg ccgactacgt gatgcgcatg gtgcacggcg   20520
acatcgcgcc cgacttccgt ccgcgcacca tcctcgattt cggctgcgcg gtgggccgcc   20580
tcgtcatccc cttcgccgcc caggccgagc aggtgaccgg cgccgacgtc tcccccctcga   20640
tgctggccga ggctgcgcgc aactgcgccg agcagggcgt ggccaatgcg cgcctcgtgg   20700
tctcggacga ctcgctcacc gggctgccgg gccccttcga cctcgtgcac tcgttcatcg   20760
tgttccagca catcgatccg gcgcgcgggc gcgacatctt ccgccgcctg ctgggcaccc   20820
tggcgagcgg cggcgccaccg gcgttgcact tcgtctatgc gaagcgtatc tacgcggcga   20880
cgtacggcgt ggcgccaccg cccgagccgc cgccgcccccc gccgccgccg cagcctccac   20940
ccagccgggc ggagatcaag gccgcgagca aggcgcgggc cgcgctggcc acgcaggcac   21000
gggcgccgga accggcaccc aatcccgacc ccgacatgca gatgaattcc tacccggcgg   21060
cggagatgct gttcctcgtg caggaagccg gcgtcacgcg cttccacgtc gagttcaccg   21120
accacggcgg cgagctgggt ctcttcctct tcttccgcaa gccctaggcg tccttcgagg   21180
cgaaggagag gcccagcgcg gtgatcgtcg cggcgtgctg cggggcgacc gccgccgcgg   21240
ccgcgacttc ggcagccacg cgctcgaggt cgccgcgcga cccgtacagg cgcacgcgcc   21300
agtagtgcac ccagtaggag tccggacagg ccgacgcgat tcccagcgcg cgctcgacct   21360
cgggcccgac gcgccctgg tgcacttggt gcgaggccca ttgcaggttc acggagggct   21420
cgaggggcag gacggcgagc agcgcctcgc tcaccgccag cgcgcgcgcc tcatcgccgt   21480
cgtcggcgag gcgcatcagc gtgggcaggg actccggcgc aagcgcgtgc gcggccgcga   21540
gcgaatcgcg cagcatcgca gcgtcgccgc gcgcgagcgc caggcgggcg cggtggtagt   21600
gggtccagaa cgaaggcgcc gcggcatcga gcgcgcgcgc gacctcggcg cttgcgtcgc   21660
ccccggagag ccgggacgcg gcccacaaca ggttcgcctc cgccgcgtgc gccggcacgc   21720
tccgcaacgg cgccagcacc gcggtggcga cctcgcgcga accctcgcgc gacgagaagg   21780
ccgagcgcgc gaactgcatc agcgattcgc cgagcgacga atgttcggcg cagagccgcg   21840
cgggatccac cgcgccgagc gcgcgggcgg cacgcagcgc cgcctcctcg tcaccgcgaa   21900
gcgcgtgcag ccgcaggcga tggaacgcgg cccactccgc cgacatcgac gccgccccgc   21960
gctccagcgc accctcgacg gcgccatcgg tgcgcccgcg cgcgatgtgc aggatggccc   22020
aggtcgcctg cgcgccgggg aggttcgtgc cttgcgcgac gagcgggtcg aggaccgcga   22080
gcgccgcggc ggcgtcgcgc gcgccactga agtcgttggc caggccgatc agcgcccccgg   22140
```

```
cgagctcggg ggccgcgcgg aaggcggccg ccgcgtcgcg ccccaccgcc gcgaggtcgc    22200
cgcgcagccg gtagaggcgc gcgcggtggt aatagcccca gaacgggtcg gggctcgcga    22260
tcagtgaatc gaacgagcgc tcgacctcgc cgcccacgtt gccgaggtcc acctggtgcg    22320
cgccccacag ggtcgccgca tgcgcggcca tcggctcgtc gggggacgaag gcgcggtgga    22380
tcagctcgac gtcgccgaag cgccgctcgc gcaacgcgag ctccagggcg ttcagccgct    22440
cgaggctcgc cttgtactcg cgataggcat cgcggtcgcg gtcgtattcc tgcgcggcac    22500
tgtcgtccat cagcgaggcc gtgagccgct ggtcgccgtg aacggtgtcg tatgcgtagg    22560
ccgcgcgatc gtcgtagggc agcgtggccg cgcgcagcga cccgaacgcg gcgcggtcgc    22620
ggaatcccag cttgcccagc acgcggggtg ccaggtcgtg ggcttccatc tcgcccgggc    22680
tccagtcgcg cagcgccggc acgttcgagt agagccacaa cggcgtgcgg tagagctgcg    22740
ccaggtccgc gtccgggcgg ccctgcgcct cgaagatcgc cttccactcc gcgagtccct    22800
cgtggccggc ggagaagatc ggcacgtggt cgccgaagat cgccagcacg tagggccgcc    22860
cgcgcacctc gagacgctcg accagcgagc gcagcgcgcg gtcggtttcg gccatcaccg    22920
cgcagtactt gtcgatcgcg ccctggcgca cgtgcccggc gcccatcagc gagacgccga    22980
agcgcggctc ggcgcccttg tcgcccaggt agggtccgtg ggcatccatc acggccgaat    23040
acacgaagcg cggccgcgaa tcgtcgagct ccgcgatcac ctcgtcggcg tggaacgccg    23100
ccgcgatgcg ctcgaactgg cgcggcgcat gggcgaacgc ctcgtcggcg tggaacgccg    23160
cgaagccgag cttcttgtac acgcggtcgc ggttgtagaa ccagcggtgg aacggatgga    23220
tggccacgct gcggtagccg tggcggcgca gcatgcgcgg caccgcatac acccggtcgc    23280
gcaactgccc catgtagggc agcgcgcggc agccgatcac cttcacgcgc aggcggccgg    23340
ggatctcgaa ctcggtgtcg caggtgttgc cgcggcagcgg gtcgcgcgcg agcggcacgc    23400
tggtgcggcc cggccgcgtc gcgccgagga acggcagcgg gtcgcgcgcg agcggcacgc    23460
cgaacgggcg caggtcggcg agcgactcca tcatcaggaa gatgatgtcg gggcgctccg    23520
cggtggccgg gccgctcgcc gagcgctcct ggcgctccag cagccgctcc acgccctcgc    23580
gcagccgccg cggcgcgacc gggcgggcga tgctcccggc gaccagcacg gcgtgcagga    23640
cgaagcccag gcgccgcacg ttcacgttgt ggtcggtggg cgaatacgtc atcgcgtcga    23700
actccaccgg ccagcggcgc aagcgtccgg cgacgaaggc gacgccgagc gcggcgagca    23760
cgagcagcca ctgccggcga tcgtcgtagca ctgcggcagc ggcaagcgcg cgcaccgcca    23820
cgaacgcggc ccacgccagc gccgcggcga acaccagcgc ctcgcgagcg ctcaggaaat    23880
cgcgcgggcc atgcaggtac gccggcgcga acaccagcgc ctcgcgagcg ctcaggaaat    23940
cgcccgggcc gaccacgcgg ccgaagaacg cgagcaggcc tgcgttcacc gcgtagagcg    24000
cgcccgccag cgccacggtc ccaccgtcg cgagcgcccc gaccgcgacg aagagcagcg    24060
catggaagat ggcgagcagg ccggcgttcc aggccaggcg ccgcgcgcgc gccacgggtt    24120
ccacgtcata gcccaccacg aggcgccgct cgagggcggc ctgcagcacc gcgtaggcga    24180
gcacgcagac cagcgtgccg cccagcagca cgccgaggaa acccgccaaa ccagccgtgt    24240
cgtcgaggat catgggacca ccaggacgt cttccggcca gtctataacg ccccgcccca    24300
tgaaaaacgg cgaggccagc gccggggccg atcaggtttg cgctgccgat ggtggcgagc    24360
gggggtgcga ccgaccagcc gccggggccg atcaggtttg cgctgccgat ggtggcgagc    24420
ccgttcatca ggcgcaccgt gacgctgccg tcggtgtggc ggaagaccag gtccttgcgg    24480
ctgtcgccgt tcaggtccag cagctgcgtc acgttccaac ccgtgcccgc cggcagcgca    24540
tcgcccgcgg tgatgatgga aacaccatcc atcaggcgga tgtgcccagag gttcaccgac    24600
tggcggaaga ccaggtcgtc cttcgcgtcg ccgttcaggt cgcccaccag gttcaccgac    24660
caccgctgc cggccggcag gagctgggcg ctggcgccga aggcgcgtcc gtcgccgttg    24720
aacaggtgcg cgcggccgtc ggtgtggcgg aagatcatgt ccgcgcgtcc gtcgccgttg    24780
aggtcgccca ggtggcttac cgtccagccg ctgcggccg agaggaagcc cgctccggcc    24840
gtcaccgtgg tgccgtccat gatgtagatg tagccgcggc cgtcgctgtg catgaagacc    24900
aggtccgcct tgccgtcgcc gttcatgtcg cccgcctggg tgagcgtcca gccggtcccc    24960
gccccgaaga gctgggcgct gccgatgatc gtcgtgccgt cgccgttcag gtccgccgtg    25020
cgtccgtcgg tgtggcgcag gatgatgtcc gccttgccgt tgccggccgt caccgtagtc    25080
tggctcacga tccagccgag cccggcgggc agcagctcct tgccggccgt caccgtagtc    25140
ccgttcatga tgtacgcgta ggcgcggcca tcggtccagc ccgcgccggc gccgatcagg    25200
aggtcgccat tgaggtcggc cgtctggctc accgtccaca ccgcgatgcg cccgtcgctg    25260
ttggccgagc cggtgatggc ggtgccgttc atcgtccaca ccgagacgac cggcttggtg    25320
ttctgcagca cgaggtcgct cctgccgtcg ccgagacgac cggcttggtg    25380
ggcggcgtga cgccaccgga aaccggcagt gccagccgt cgatccaggc gcggtcggcg    25440
cagttggcgc catccggggc cggcgggttc caggtcggcg agttgcaagg cgtggagagg    25500
aagttcgaga aacgccacac cagcgtgtgc gcgcccgcgg tcaccgggaa cgacaccagc    25560
```

```
ttccagcccg aggccgtggt gcccgcgtcg ctgaacacca ccgtgccgtc gatcaggaac    25620
tcgaacttgc cggcgttggg gaagctcgac acgcgatagg cgaacgcgac gttgcccgcg    25680
agcagcgtgc ccgcgaacga gaggtcggag ttcaccgtcg tcgagttggt cgggtcgctc    25740
gtcaccacct gcgccgaacg aaggctggtc gcgccctcga aggcctggtc ggagccgacc    25800
gtccatgcgt tggcgccgcc cgaggtcgtg aagccggcag gcagcgtgcc acccgtgggc    25860
cacggcgaga ggatgatgcc ggtcgcggtc gccgggccgc cgatgaagac acccgtggaa    25920
cccgtcgcat tggagagcgc gacgttgaac gtctcgttgc cttcaccgat accgtcgctc    25980
gccaccggga ccgtgatggt cttgggcgtg gtctcgccgt tggcccagct gagcgagccc    26040
gaggtcgcgg tgtagtcggc gcccgaagtg gcggtgccgt tggaggtggc gtagctgacc    26100
gagatggcac ccgccgagcc gccgatgcgg ctcacggtga gcgtcacgtt gccggcggtt    26160
tccgcggccg cgaaggtggt gccggtgaac tgcaccgaac cgggaacggc gggcggggtg    26220
accgtggact tgagcgccga gagtgccgcg cggttgttgt tcagcgagag cgcgtcgttg    26280
gcctgcgcgg tgccgcacgg ttgcggcgca ccgacacccg tcgtgcacga gaggcccggg    26340
ttggagaagc ggtagacctt ggtcgtcgag tcatgcagca gcgacatgat ggtggtgaag    26400
ttcgccgcgt cgttcgtcgt gcactcgggg cgctggctca cgccgcaggc gccgccgccg    26460
gagccgttgg gcagggccga gttgcaggtg aggttgccgc tcgcgcaata gaagtggccg    26520
tacgagtacg cgaacgcacc ttgcggcggg ctcgcggtgc cgcccgcttc ccaggcctgc    26580
gaggggcggt cgtggcggtt gcccatggca tggccgagct cgtggatgaa gacccactcg    26640
caaccgaaca ggcagccgga agcaacggcg taggcaaagt tggggtccgg cgtgctggtg    26700
ccgatccagg ctgcgccgtt gccgccgaag tcggcgccgt tgcgcatgaa ggccaccatg    26760
tccgcgcccgt attgcgtgcg gatgcgcctcg atgttgccga acgtcgcggc gtcgaagctc    26820
gcgtcgcccg gggtgatggc atgcatcgcc gtggtgtcgt cgatggcgtc gtcgtagttc    26880
acctgcgtcg cattcaccag gcgcagcgtg atggcgacct cgctgtcggc atacgcagtg    26940
ttggcgcgcg tgacgaggaa gttcaggcgc gtcatcaggc tgccgccgag gcgctggccg    27000
aagccgcgcg agtacacgat catcaggtcg atgacgttct gcggcgtggg cgtcacctgg    27060
gcgagccgac ggtattcacg ccggggatgg cgatcagggt ctcggggctg ccggtcgcga    27120
cgacgcgcga cttcgagccg gtcgtggcca cggcgtggtc ggcgccgagg ctgatgatcg    27180
gcatgctggc ctgctcggcc gtcatgtcga ccaggtagtc gacggtgccg cccggaatca    27240
ggcggtactc accctgcggg gtcgagaaca cgccgaagga gccttcaggc ccggtcgtga    27300
cgatcgcgcg gtggttgatg ccgtcggtct tgctcttcgc gatccacgac gtgatgccgt    27360
cgccgtgggc ctggagcaac tcgaagacat acgagtggcg cacccgttg gggagcgaca     27420
gctcgacctc ggagcgcgga ctcagggcat ggagggccgc ggcgttgaag cgcaccgctt    27480
gccgggcaac ggcgctgccc ggcgccggca tcgcggcggc gggtgcggag accaggatct    27540
ggggaaccgc ggcgaaggcc acggaagaga agacgatcgc cagcgcgccg aggaaggcgg    27600
ctgtgatccg ttgcgtgaat gcgttcatgt ggagctccgg aagttgaccc atgcccaatc    27660
cgctatgtcg cggagatgtg dacaaaggt atcgagcggg cgtgacgacc cgccccgga     27720
gggatgctcc aaaaggacta cgagggtgct acggctgggt tggtggcgaa ggccgaagga    27780
gaactccttg tggtctccgc gcgacttaag gttgcgaggg aaccaccgac cacccaccgg    27840
ggccgatcaa gtccgcgctg ccgatcgtgg tgacgccgtt catcaggcga atcgtcacgc    27900
gcccatcgtt gtggtggaag accacgtcca tgcgcccgtc gccgttgaag tcgaggagct    27960
gcgtcaccgt ccagcccgcg aacggcggca ggatgtcggc ggcgccgagg atcgcggtgc    28020
cgttcatcag gcgcacgtgc gcgcggccgt cggtgtgcgc gaacacgagg tccgcccggc    28080
cgtcgccgtt gagatctccc accaggttca ccgaccatcc ggtgcccgcg ggcaacagct    28140
cggcgctggc gccgaacgtc gtgccatcca tcaggaacag gtgtgcgcgg ccgtcgttgt    28200
ggcggaagac gaggtcggcc ttctggtcgc cgctgaagtc cgcgacgtgg ctcaccgccc    28260
agccgctgcc gggcgcgagg aagcccacgc cggcggtgat cgccgtgccg ttcatgatgt    28320
atgcgtagcc gcgcccatcg acgttggaga agacgatgtc ggccttgccg tcgccgttca    28380
tgtcgccggt gccgaccacg ttccatcccg agcccgcggg cagcaggctc gcactccgc     28440
tgatcgcggt gccattcatc agccagatgt gggcgcgtcc gtccgtgtgc tgagcagcag    28500
gtcggccttg -ccgtcgccgt tcaggtccgc cgtgtggctg atggtccagc ccgtgccggc    28560
ggggaacagc tccttgccgc ccacgaccgt gagccgttc atctggtacg tgtagatgcg     28620
cccgtcggtg tgctggaagt agatgtcggc ctggctgtcg ccgttggaat ccgcggccat    28680
tgatcacgct ccagccggcg ccggcgggaa tcaggttcgc cgaaccagtg atcgtcgttc    28740
cgttcatggt ccacgcggcg atgcgcccat cggtgtgctg gaacaggaag tcgctctgcc    28800
cgtcgccgtt caggtcggtc gccgcatggg tcgggaacgc caccttcgcg aggaaggcat    28860
cggagaaacc cgtgagcaac gtcttgtacg cgccgggcgt ggtcggatag ttcgccgacg    28920
acgtgcggcc ggccacatag gccgcgcccc gcgcatcgac gccgacgccg taccccacct    28980
```

```
cgacgttgtt gccgccgatc agcgtcgaga acacgatgtc gccgttgcgc acgcgagtga    29040
ggaacgcgtc ggcatcgccg ccggggccct gcaccgggtt cagggtcggg aagccggcga    29100
cctgcgtgta gcccacgacg agcgcatcgc cattcggcgc cagggcgacg tccaggggat    29160
tgtcgtaggc cgagccgccg aagatgcccg acgattcgag cgcgccggtg gccgagaagc    29220
gcgtcacgaa cgcatcgacg gtgccggcga aattgcgcac cgttccggtc tgcgggaagc    29280
tgggcaggcc ggtatcgccg acgacggtcg cctggcccgt gaccgggttc accgcgatcg    29340
cggtcacgtt gtcgccgctg ccgccgagga acgtggaata cgcgaccggg ccgccgctcg    29400
caccgatctt ggtgacgaag ccgccaccgg gttgggtctg gtacgcattc acgttgcgga    29460
acgtcgagga gcgcgtggag cccgcgacgt acgtgttgcc cgcattgtcc gttgcgatac    29520
cgcggccgat gtcatcaccg ctgccgccga gatacgtgga atagacgaag ccgccgttga    29580
acgagagctg gtgacgaacg cgtcgcgatt gaggccggtg ccgacgaaca cggctgcagc    29640
gcgctgccgc cggtcgggaa gccgcgcgcg gccgaacccg tgacatgcat gaagccgttg    29700
ttgtccacgg cgatcgccat gccctcgtcg tcgccggagc cgcccagcag ccgcgagaag    29760
atgacgttgc cagtggtatt cagcgccgtg acgaatgcat cgcgcccgtg ggcatggttg    29820
gtgcaatcgg tgacggcgaa gcaatgcggg aacgcgttgg tgccggtgga accggtgatc    29880
gcgacgaact cgttgccgcc ggaggcgccg agcgccagcc cgcggagata cgcttcggcc    29940
ccgagcgtgg tcgagtacac cacctgctgc gtggtggggt tgaacttcac caccgtggct    30000
tcgtacgtgc ctgcggtata ggtcgtcgagc ccgacgtaca cgttgccctg cgcgtccacc    30060
ttgaccggcg tggtccactg ctcgtcgccg cgcgcacgca ggtacgtgga gaacgccatc    30120
accgggtcga tcaccagcgt gcgggtgctg tcgtagtcgg cgatcacgaa gccggcctcg    30180
gcgccctgct cgcccacgaa gagctcgaag cgcgcggcca ccggcacgcg cgaatcgccg    30240
acctgctgga aggctaccgg cgcgtgctgc gtaaattcct cctggcccac gcgcatgcgc    30300
aggttgccct cgccatcgat ccacgcgtgg tcggcggacg agaggtcgag gcggatctgg    30360
cgcggatcgg cgcgcggggc gaccacgaag tcgtattcga gcgtgccctc cttgccgtac    30420
acggtgaggt ccacgccgcg atacaggtcc ttcagcgaca cgcggccgaa gtgcggcacg    30480
ttttcgcgcc ggccggcggc cgtggtgccg ctgtaatagt ggctgagggt ggcccggggc    30540
tcctcggcct cgatgaccgg atcggaagcc cccgcgaagc gcacgcgcaa cagcgaggcg    30600
ctggcagctt tcctggagcg tggctcgaag gcaatgccat cgcggctcac ggagacgcgt    30660
ccggcctgcc cgcgcgagac gtagagcgca tccgggccga actggccctc gttgcgctcg    30720
aacgtgatgg gcgcgttcct gagggcatca gggacgacgg catggacggc ggcgggagtg    30780
gcgaggcgg ccgcgacgag cacggccagg gggcggagcg ccgggcgggc gccgctggag    30840
ctgcgggtca tggggcagtg acctttctgt tgttatggct gcttgttgtc tttgcgggat    30900
ccacagctcg gatccagctg gcgatactac cagcgagcga gtggttttc gtgacccagt    30960
cccgcagtcg ggccccggcg acaacggcgg catcgcggtc ctttgccagg gcacgcacgg    31020
cctcgatcca gtcgccgtcg cgggccagca cggcgggggc atcgcggtag ggctcgagat    31080
ccgaggccac caccgggatg ccgaggcagc cgtactcgag cagcttcagg ttgctcttcg    31140
cgcgattgaa cgggttgtcg cgcagcggtg ccaccgcgac gtcgagtgcc agcgacgcga    31200
gcttctccgg gtactgcgcg atgggcacca tgtcgtgcac ttccgcggcg aacggcgcca    31260
gctcgggcgt gcacaggccg aggaacaccc agtcgatctc gcgatgcgtg gcgcgcacca    31320
cgggctcgag caacttcagg tcttccccat gctgtttcgc ccccgcccag cccacgcgcg    31380
gacgcgcgcc acccgcggga cggttcgcga ggccggccca acgctccgca tcgatcgcgt    31440
tcgggatcac gcgcacgtcc ttcgcgccgc ggccgaaggc ctccgcgagc ggtgcggtgg    31500
acaccacgag gcggtcgcac aacgccaccg cgcgcgcgat gcgctgggcg atgtccgggt    31560
agatcgtcgc cgcatacgga ttgccgggcg gcagttgcgt gagcaggtca tccaggccca    31620
ggaccttcag cgcgcgtccg tggcgcgcga gcacctcgag cgaggtgagc tggtagttgt    31680
ggaagaagtt gtgcgcgagc acggcgtcgg cgtcgaggcg ctgccactcg acgcggttgg    31740
gcgcgcagcc cgtggcgtgc tcgcccatca tcacgacctc gacgcggccc gcgcgctcca    31800
gcgccgcgca cggctggcgc acgcgcacct cgccggagcc ccagcggtcg aacggaaacg    31860
cgcagagccg aaccgcgtgc gtgtcgttgc cgcgcggcgc gaagcgttcc acgggcccgt    31920
agccctcgcc gcccatcgcc aacgccgggt ggtagtgcgg atcgtcgtcg agcaccggct    31980
gccagcgctc gcgcatccac gcctcttcgg aaggaatgac gagcttcgcc accggaacga    32040
cggcatcggc gcgcgcatcg gcgagcaacg ggaagtccgc cgcgatgccg gggcgcgcaa    32100
ggatgtcgaa gcccgctgcg cgcagcccga ggcacaggtg cgccatcgcg aacggccccg    32160
cgcgctcgat ctcgtgcagc gcctgcgcgg agagcgccgc gtcgcggttg accagcgcga    32220
gacggcccgc cgcgacgctc acctcgcgcg gcgatgcgta gagtgccgcg aggcgctcct    32280
cgccctgcgc gcgcggcgga gcgcccgaga cggcccacgg cccgccgccc gcgatttccc    32340
agccgggaac gcgcacaccg ccggggccca gcaggtcggg cgcgagggca cccgtggagg    32400
```

```
ggcctgcgat gccctgctcg aggcgctcga gccagccggc gcggaattcg ctgcaacgcg    32460
catcgacgat cgcgatccac agcgtcgccg cgttgcggat ggcgcgcgcg aacgcggagc    32520
cggtttcgcc gcccaccgga accaggtgcg tgcgcacgcc gcgcccgccc agctcaccgg    32580
gccagcgttt cggcgccctt cgcgtcgaga tcgacgtaca cgtggccgat acgttcgcgg    32640
cagcccttca cgagcgcctc gatgcacgcc accagccgct ccgtcgcgtg cgtcgcgtca    32700
atgaaggcgc aaatcgtgcc ggccccgatc gcgcgcgccc acgatgcggg cgcatcggac    32760
gcagctggca ggaacgcggc ctgcgtgcgc gtgcggcgga agtgctccgc gaccacgcgc    32820
tggtggcgct cgcgcgcggc gtcgctttcc tgcgcggcga tgttcaggtc gaggcgatgc    32880
acgagcagct cgcgcagtgc caccacggcg ctcgcacggc cctcctccgc gaggcgcagc    32940
gcgagatcgg tgacgcccgc ccagccgccg gcgcgcagcc cgccacgctc cagcaacgca    33000
gcgcggcgca gcaccgcgag cccggcgagc ggctggtgtc gcgcgaactc cgcgccgaac    33060
gtggggacga acgccggatc gtagcgttcg ccgtgcgcgc tcagggcatc actgcagccg    33120
tagagaagct cggcgttcgg gttgcgtgcg accgcacccg ccaccgcgct cggcgcgcca    33180
tcggccagcc ggtcgccctc ctccaccagc acgacccatt cgcccttcgc gtccgcgatc    33240
gcgcggttga cgcgcccgag cgccgcgtcg ctcgcgtcgc gctcgaggac cacggtctcc    33300
agggcgatcg cggcagtgct cctgccgcgc aaggcctcgc gcgccatcca gtggtcgtag    33360
gcatcgccgc ccagcggcgt gcgctcgcgg cggcgtggaa agccgaggca cgccgcggcg    33420
aacaacggat cgatcgttgc gggcacagca tccaccgcgc catcgcgctc cgccgccagc    33480
tcgcgataga gcgcgcaggta ctcggccgca ttcgcctcga tcgtcttcgg gcggtggtac    33540
ttcaggttct ccgcgacgcg atcgagcgcc gtgcggtcgt cgcgcaggcg cagcaggagg    33600
gccacgatcg cgccgggatc gttgtgcggc agcaggaagc cggtctcgcc atcgcgcacg    33660
cgctcgggaa tggcgccatc gcgcgtgccc gccaccggga tgccgaggcg ctgtgcctcg    33720
ctcaacacca ggcagaaggt ctcctccatc gtcgagggca gcaggaccag gtcgatgccg    33780
cgaaggatgc gcggcaggtc cgtcgcggca taggcgccgc gctgcatgat gcccgccgca    33840
ttcgccgccg cttcgagcgc gagatcgcgc ggaccccagg cctcgatcgc gatgcgctgc    33900
cccgccagcc ggcgcgcggc ctcgatgatc aggtgcgcgc ccttgcgccc gccgaagcgg    33960
ccgaggaagg ccacgcgcag caccggccgc gcttcggggc gcggaagcag cggcaggtcg    34020
gcaatgccgt gcccgatgac gcgcagcttc gcttcgcagc tcgcgccgaa gcccttgcgc    34080
atgcgcatcg ccgcgtactc ggacgggcag aggatcacgt cggccgcgtc cacggcatcg    34140
gacgccgcgg cgaagcgcga ttcgaggtag ccggtcacgg acgccggcgc cgtcggcccc    34200
acgatcgcct cgctgcggcc gcgcaggcaa tccacgcatc cgccatcggc gccgcgcgcg    34260
gcggcgcggc cgcacggcag gtcctgcggc ccggtcatca tgttgtagtc ggcgcacagg    34320
tacgagaagt cgtgcgccgt gatgaccacg cgtgcccccg actcacgcgc gatgcgcggc    34380
aggcgcaggc tgttccagcc caccaacgac tggaagtgca cgacgtcgta gcccccggac    34440
agcagttcgc ggaaggccgc ctccacgccc gcgtcggcaa ccgccgcctc gtgcccgatc    34500
acgcggatgc cgggcgtcat gcgctcgggg ttcaggcgcg cgagtcgcac gtgcgtcgcc    34560
aggcgctcgg tggtgaagtc ggtccacgcg cgcgggccct cgggcagcac cagcgtcgag    34620
gcgacgtcct tgcgcagcag gttcaccagc gcgcgcgtgt gctcctcaat gccgccgcgc    34680
gagtcgaggc gatgcagcac ctgcagcacg cgcgggcggc ccggcatccg ctcgcgctcg    34740
gccgcggcgt tgacgcgctc ggcggcgcgc cgcagcgggt tcgcgatcgc ccacgcgcgc    34800
acgcccgcga cgtaggtggg ccagcggcgc tcgaggcgca cgcggttggc ctcgcgcgcg    34860
gcgtcgatgc cggccacccc gccgaacgag gcctcgcccg cgtggtgcac gtagacgtcg    34920
tcgcagcacg cgatctgcca gtcccggtcc agcgcgcgca tcgagaggtc gttttcctcg    34980
ccgtagccgc ggccatagag cgagtcgaac gggccgcatt gcacccacac ctcgcggcgc    35040
accagcatgc agaagccgac cgccgtcggc aggcgcggat actcgcggcg gctggcgcgc    35100
gcgacgcagg cggagatggc atcgacgtcc ggctcgccgt cgggccgctc gaagatatcg    35160
gcgacgccgg gcaggctgag gatcgtcgcg cggttggaga gcgggcacac gatgccgacc    35220
cgcgcatcgc tgtcgcggca acgcagcagg ccctcgagcc aaccggggcc gacctgcgta    35280
tccgaattga gcaccacgaa gtcgccgccg ctgccggcca ggcccgcagc gaccgcgcct    35340
gcgaagccga ggttcttcgc attggtgacg acacgagtgc cggcacgacg cgccgcgaag    35400
ctgcccagca acccggcgag gcgcggatcg gtgctcgcgt catcgacgag gacgatgccg    35460
tgcgcggcgc ccgtgtggcg ggccagcgac gcgaggcagc gttcggtgtc gtcgtggccg    35520
ttgaagaccg ggacgacgac tagcgccggc gccatccgac cggcgcttcg agctgggcca    35580
atcgcgccag caggcgctgg tgctcctgcg agagtgcctc gagctgcttc tgcgaggcga    35640
cgtaggcgcc tccatgcgct cccaggtgcg tgcgcgcctc ctcgagctgc cgcagcgact    35700
gctcgtggcg accggagacg tcgtggtcgc ggtcggagag cagcgagaag cgcccggccc    35760
ccgccgtcac cgcctcgcgc tcgccgcaca gggcgaggaa gtacatcgca tcggcgacgc    35820
```

```
ccgccgcggg tttcgcagcc tccggcgccg aagcttgcag caactgcacg gcgtcgggca    35880
ccgcgtgcag cggccagatc gccgaatacg catcgacgcg ttgcccgaac agggccacct    35940
ccgggaaccg cgccttcagc acggcgagga attcgtgctc gtagagctcg cgcacgtgat    36000
gcgggttgcg gtagtcgcgc tggtccgagt acacctcgcg gttcggcgtc gagaccagca    36060
gcaggccgcc gggcgccagc acgcgcttcg cttcgtcgag caggcgctcg ggatcgggaa    36120
tgtgctcgag cgtctcgaag gagacgagga gatcgatgct tgcgtcatcg cacggcagcg    36180
actcgcagcg gccctcgacg tactgcaggt tctgcgcggc accatagcgg cggcgcgcct    36240
gcgcgatggt ctcggcggcg acatccgcgc ccaccacgga cttcgcgcgc gtcgccagca    36300
gcgccgagcc gtagccctcc ccgcaggcga tgtcgaggac gcggcagcct ccggcgagcg    36360
gcaacgcaaa gtggtagcgg tgccagtgct cgtaccagat ctcgccctcg aagccgggct    36420
ggaaacgttc gtgttccatc gggagagtat aggtggatgt gcaaccgccc tccggagagg    36480
gcggttgcga tgatcttatt cgagatcggc tgctacggaa gcaccggcaa cggctgcggc    36540
ggcgccaccg accatccggt gcccgcgccc agcaggttcg cgctgccgag cgtggtgagg    36600
ccgtccatca ggcgcaccgt gatgcggcca tcgacattct tgaacaccat gtcgagcttg    36660
ccgtcgccat cgaagtcgag cagttgcgtg acggtccagc ccgccccgc cggcagcacg      36720
tccgccgcgc tgaggatggc ggtgccgttc atcaggcgca cgtgcgcacg gccgtcggtg    36780
tggcggaaca cgatgtcacc gcgcccgtcg cggttcatgt cgcccacgtg gctcacgatc    36840
cagccggtgc ccgcgccgag gagctcggaa ccggcgacca acgcggtgcc gttcatgatg    36900
aagagatgcg cgcggccatc ggtgtggcgg aagaacatgt ccgccttgcc gtcgccgctc    36960
acgtcgccca ggtgcgtcac cgtccagccg ctggcgggcg agaggaagcc cgcgccggcg    37020
gtgatggtgg tgccgttcat caggtagatg tagccgcggc cgtcagcgtg gatgaagacg    37080
atgtcgtccc tgccgtcgcc gttgaggtcc ccggtggcca cgactctcca gcccgtcgcc    37140
ggccccagca gctgctggct gccgatgatg gccgtgccat ccatcagcca caggtgcgag    37200
cggccgtcgg cgttgcgcag caggagcgccc gccttgccgt cgccgttcat gtcggcggtg    37260
cggtcgatgc tccagccgag gccggcgccc agcagctcct tgccgccggt caccgtgagg    37320
ccgttcatcg tgtacacgta cacgcgcccg tcggtgtggt gggatcctct agagtcgacc    37380
tgcaggcatg caagcttgag tattctatag tctcacctaa atagcttggc gtaatcatgg    37440
tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa ttcacacaac atacgagccg    37500
gaagcat                                                              37507
```
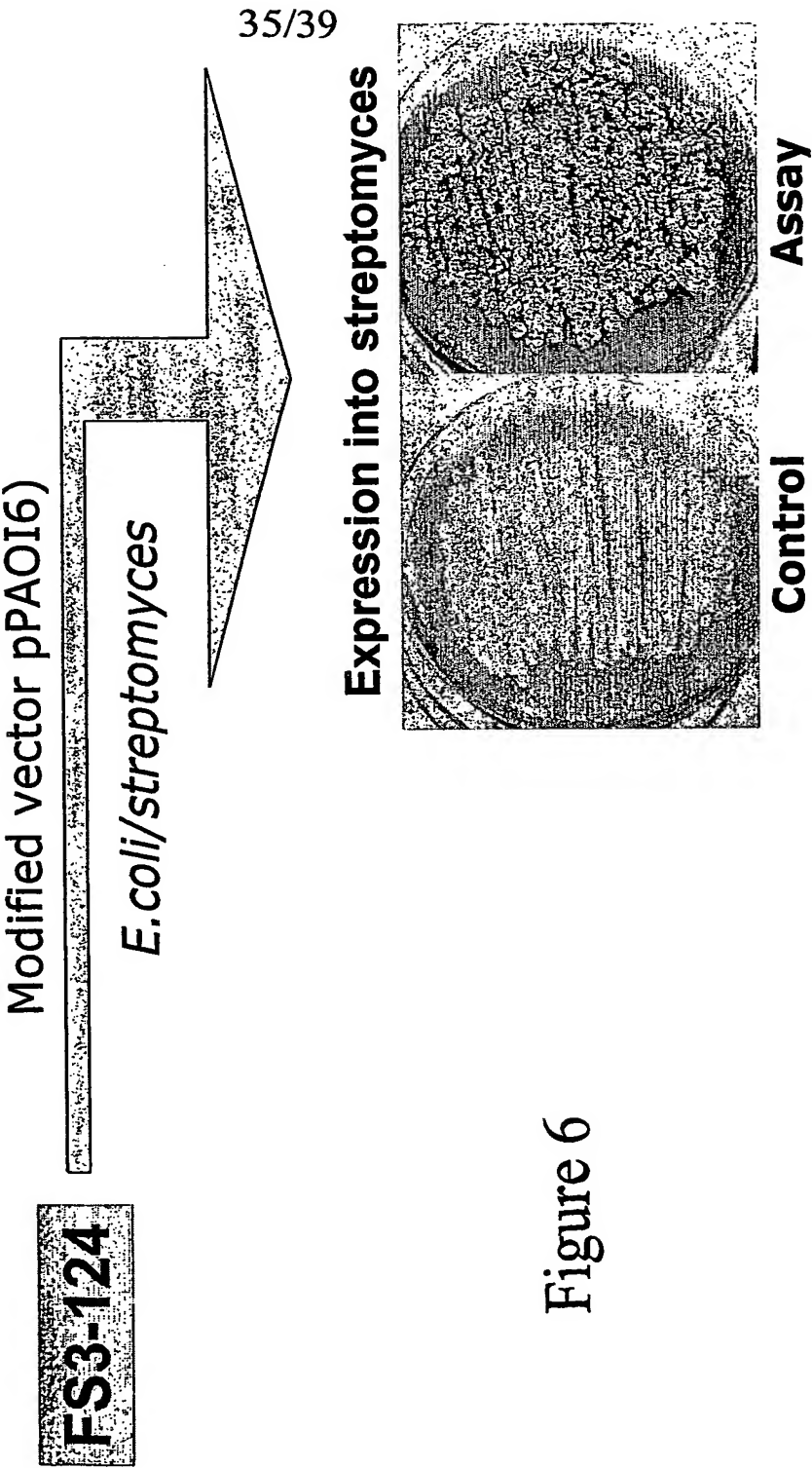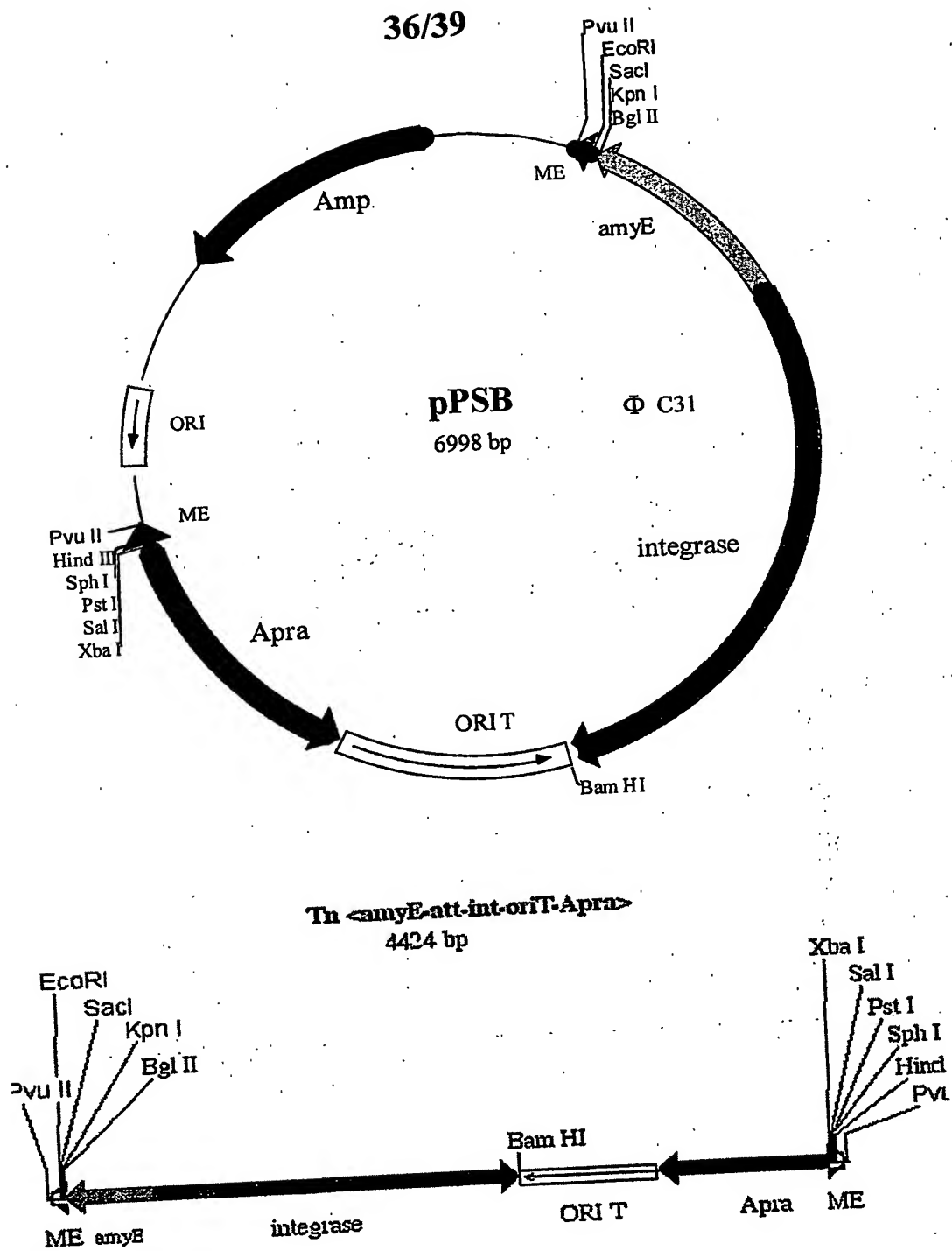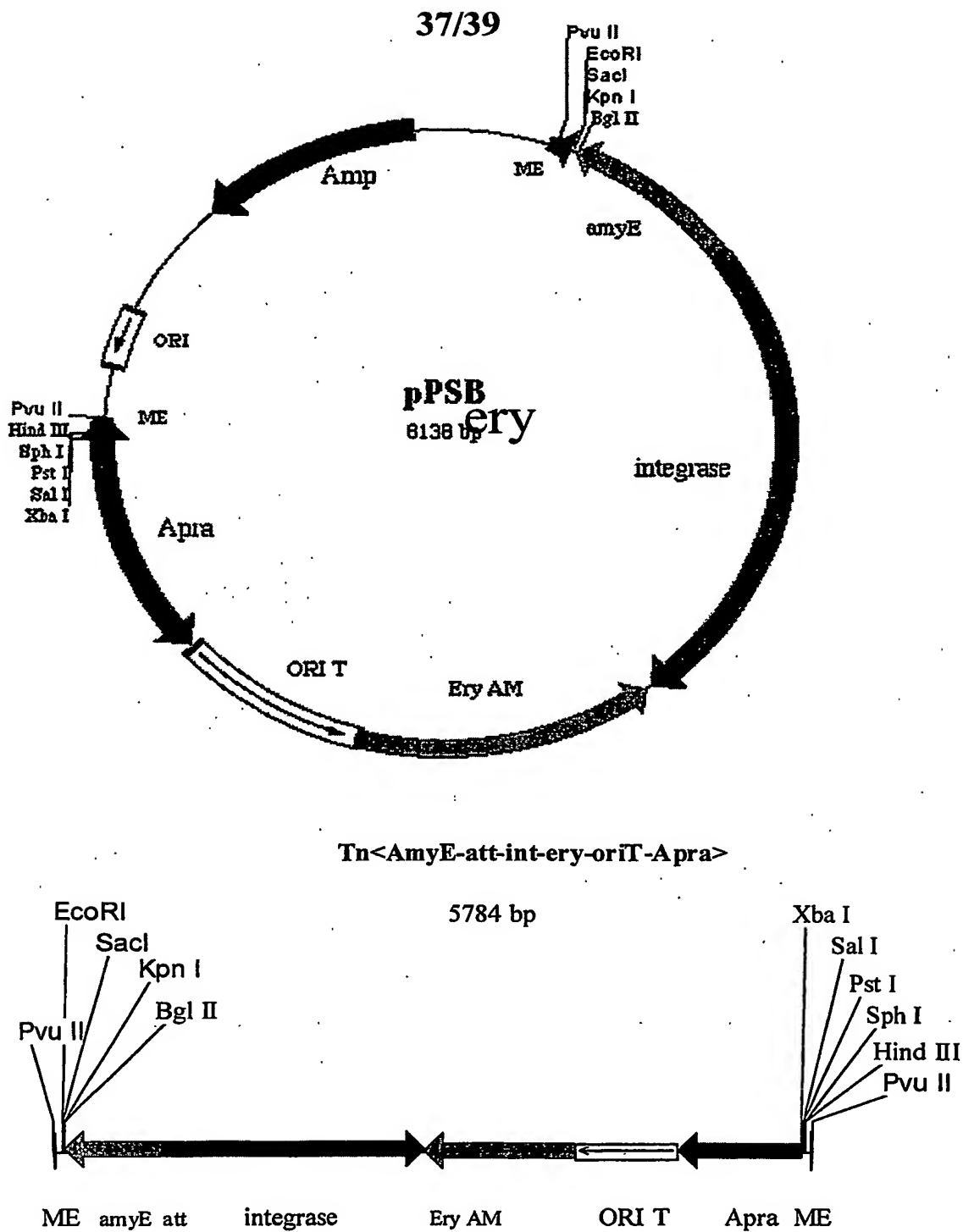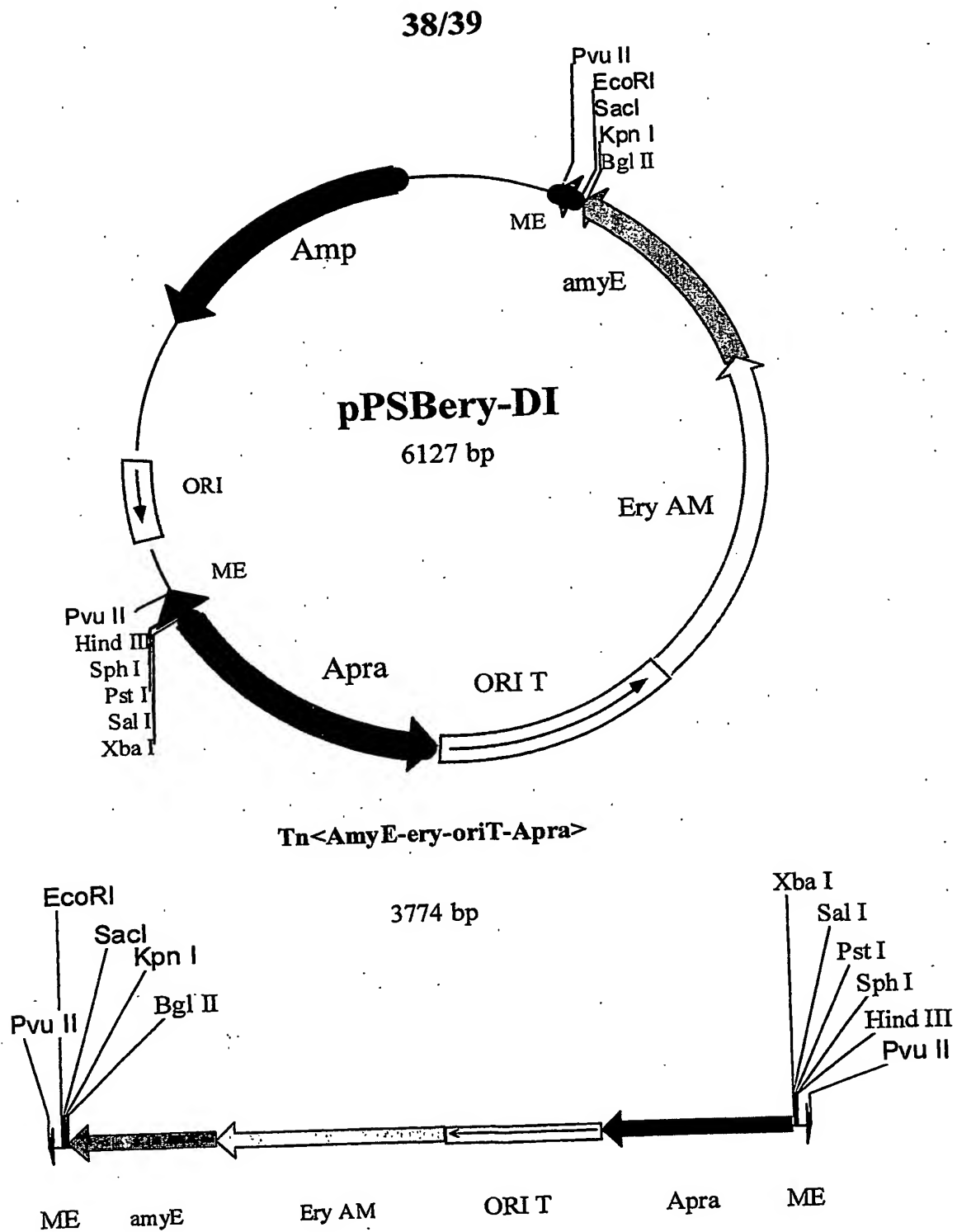
Figure 6

Figure 7

37/39

Tn<AmyE-att-int-ery-oriT-Apra>

5784 bp

ME   amyE att   integrase   Ery AM   ORI T   Apra ME
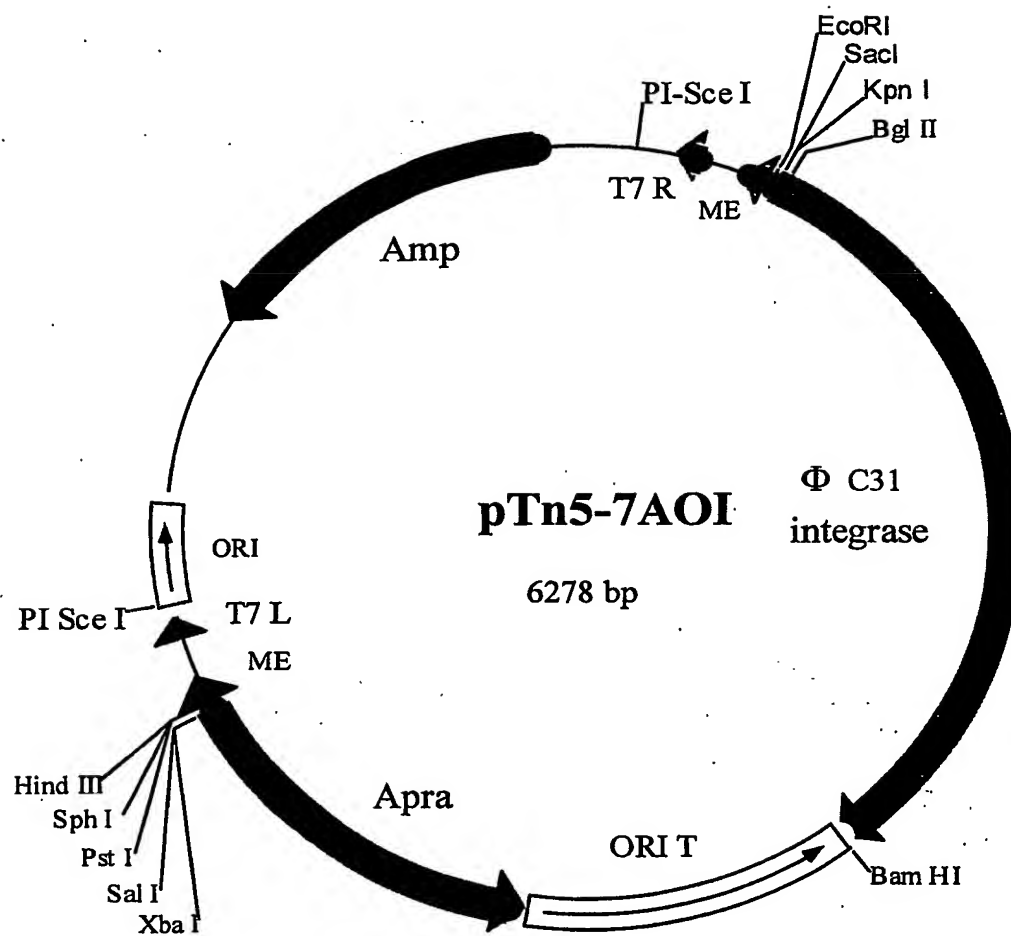
Figure 8

38/39

Figure 9

Figure 10

SEQUENCE LISTING

<110> LIBRAGEN

<120> Method for the expression of unknown environmental DNA
into adapted host cells.

<130> B0149WO

<140>
<141>

<160> 16

<170> PatentIn Ver. 2.1

<210> 1
<211> 37500
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: DNA sequence
of clone FS3-124.

<400> 1
gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa agggggatgt gctgcaaggc   60
gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg acggccagtg 120
aattgtaata cgactcacta tagggcgaat tcgagctcgg tacccgggga tcccacgtac 180
cacgagctca tctggaagag cccgctccag acacccgagg aatggaagcg cggcgtcgtg 240
agaccgtaca cgcagaagcg tctcgtcgcc ttcgttgact ggcttgcgac gacgatgaag 300
ctggatctca ccaggatgtt cgccgcaggc agctcgatgg gcggatcggg cgcgatcatg 360
ctcgcgattc gctatcccgc acgattcgcg tggaccgtgt cgtgggtcgg cgtccacgtg 420
cccgccgact ctccgcaatt cacatcgtcg tacgagctgg tgtacggccg gcccgacttg 480
aaggtgccgt tcgagaacgg cacgccggtc tgggatcatt ttagtgacgt ctggtacctg 540
cggcagcatc cggagcagga catcggggttc atcacgttct cgaacggcaa gaacgactcg 600
gcgatcggct ggcgccaggc cgtcgaattc ctgaagacgc tgcaggagac caggcaaccg 660
cacctgttcg tctggggcca ggagggacac ggccagcgcg cgaagatgcc ggaaggtggc 720
ggcgaacggg agatgccgct cgacctcaga acgaaccaga gcctgccggc gttcagccgt 780
tgcacgctcg acgacgatcc gggagacggg tcggattcga gcggcgtccc ggccgggcag 840
atcaactcct acgtcacctg gcagcctgac tcgatcgtcg acgaacccgg ccgatggagc 900
gtcgtcatca agctggcgaa tcgaccgccg cggtcgcccg ccgcagtcga cgtcacgccg 960
cggcggctgc agcagttcaa gctcaagccc ggcgatcagg tgacatggac gaactccgcg 1020
ggcaacaccg tcgtgcagcg cggcgaagcg gtcgcggacc ggtggggtct cgtcacgctc 1080
ccgcagatgc aggtctcaac gggtggaaac aggatcctgg tgagcaggaa ctaggcgggc 1140
ggcgtcctcg acgagccggc gcatcggcgc cagttgatgt gcccacgaaa gctggctgag 1200
gtccgcgcgc gcggcgatgt cccgtcgcgc ctgcggcggc tttgccagca aatccagcac 1260
agccgcccgg aacgcctcga ctgattccgc cacggagcac gccggccgtg cgatggtggg 1320

```
aagtcccct  gcgacggccg  gggtcacgac  gttcgggagc  cccgcggcaa  tcgcttcgag  1380
cactttgttc  tgcacgcctc  tcgcgatcgc  aatcggcgcg  accgcgatcg  ccgatcgcca  1440
gagatactcg  cgtacctccg  ggaccgcgcc  ggtcacttca  atcgtgggat  cggccgcgag  1500
tgcccgcacc  gcggcggtcg  ggtccatgcc  cacgagcatc  agccgcgcgt  ccgggcgggc  1560
gcgaaccacg  gacggccaga  cctctcgcgc  gaaccagagc  gccgcgcgct  cgttgggctc  1620
gtagccgaac  acgccgcaga  agacgacccg  cgcttcggcg  gccggtggtc  ctttcggcgc  1680
aaacatctcg  acgtcgatgc  cgttcggcac  gaccgacgcc  gtgaggccgc  ccgtgacttc  1740
ttcgagcagc  aggcgctcgc  gctcgctcac  gaccgtggtc  gccgccgcgc  gccgcatcgc  1800
gactttctcg  aagcggagca  agcggccggc  ctcgcggcgg  aacagccagc  ctttcggcca  1860
cgccgccgtc  gacgccagcg  cgcgccattt  ctcggagtcg  acgtccacca  tgtccaggac  1920
gaaggggatc  ccgcccagga  tcggctcgca  ggcgtagcgg  gccatcccgc  tgcagtacgc  1980
gagcaccgcg  tcgggccgcg  tcgcggcaat  ccgtcgttcg  agcacctggt  agatctccgg  2040
cgaatgcagc  aggacgtgcg  tgaggggtcg  atcgcccggc  agcgccagac  cgctcgcgat  2100
gagattccgg  ccgcggcgga  cgcgcacgac  gtcggtcgac  gccgtcacgc  cggcgagatc  2160
cgctcgatgc  gaccactcct  cgtcgtcgtg  cgcgagcgag  acgaggtgca  cgtcggcgga  2220
tcgtgcgagc  gtgtgaatga  gatgaaacgc  gcgaatgcga  tcgccccggt  tcggggcgta  2280
gggcagccgg  tgtgtgagca  gaaggactct  caccgtgcgg  catccgatcg  ctctcggatg  2340
acgcgggccg  gtacgcccgc  ggcgatcacg  cgggcgggaa  tcggcgcggt  cacgaccgac  2400
cccgccgcaa  tcacggaatc  cggaccgacg  tcagccatga  ccacggcggc  gctgccgatc  2460
caggtgccgg  cgccgatccg  caccatccgc  ggtgtgcctg  gctgctcacg  gatcggccgc  2520
gtgaggtcgc  tcgtgccgtg  cgtgtccggt  ccgctgggga  tgtggacgcc  ggcgccgacc  2580
aggacatcgc  gctccaggtg  gacccagccg  agatggcatc  cgggtccac  gtacacatgc  2640
tcgtcgagct  tcgcgccggt  ctgcgaaaag  atggtgccga  aggcgaccgt  gaccgacggg  2700
tcgcagtgtg  caagcacgcg  gccgaggaaa  gccgcgcgaa  ggtattgtcc  gatgacgccg  2760
gggatgagcg  agagccactg  cgtcgaccct  tcgagggccc  ggtcggcgcc  gaggaagggc  2820
cgccgcacgg  agtaggagag  caaggcgggc  aggacgagca  ccagcgcgat  gccacgggcc  2880
agtcccttgc  cggcgtcctt  gaggctcacg  cccactgcag  cgcgcgcccc  gcgtcggacc  2940
gggctccgta  tcgtgtgacg  agcgagtcgt  agagcgcttc  cactttgcgc  atccgccgct  3000
cgaacgataa  ctcttcttcg  atccggcgtc  gcgccgcaat  cgcacgggtg  cgtccggccg  3060
cgggattggc  catcgcctgc  tcgatcgcgt  ccgcgaggcc  tgacgcccgg  ccgggctgaa  3120
tgaccagggc  gtgcgtgccg  ttcgacacga  gctcggccgt  gcccccggcc  gcggtcgcga  3180
cgatgggtgt  ctcgaacgcc  atggcttcga  gcacggcgtt  gggcgtccct  tcgtagtccg  3240
acgactgcac  gagcaggtcc  agcccgtgat  ggacggccga  gacgtcggac  gtgtggccga  3300
ggaaacgcca  ggttccgggc  ccgagctcgc  gggcggcctg  ttcctcgagc  ggacgccgga  3360
ggctgccgtc  gccggcgatg  acgagacgga  ggcgcggcca  cttcgcctga  agggcggcac  3420
aggcttcgat  gagcaagtcg  aatcgtttct  gcggctcgag  acggccaacc  gcgccgatga  3480
cgaagtcgct  ggcctcgaat  cccagtcctt  cacgcgccgc  ggcctcagat  cgtggatcgc  3540
gtcggaaagc  atggtggtcg  atgccattcg  ggatcgtgac  gacccgatcc  ggccgtccgc  3600
ctttggccag  gagctcccgc  cgaatctgtt  cggacacggc  aatgacggca  gggaagcggg  3660
cgagaatccg  gcgatcgagc  gggtagtaga  cggtgcgttc  gcgaaacgaa  tggccggtcc  3720
atccatgggc  cgtggccagc  ggaatcacgc  gctcggcgcg  tcccaggagc  agggcgagga  3780
gatcggtctt  gtactcgtgc  gcgtgcacga  tgtcgatgcg  cagatctttc  acgagcgcgc  3840
ggagcttcga  ccagattccg  gtgtcgaacg  aatgcttctc  ccggacctcg  acgtagtcga  3900
tgccggccgc  tttcgcccgt  tcgtccatgc  cgaacaccgg  atcgcgctgg  tcgcggatgt  3960
agcagacggt  gatcgcgtac  ttggagcgat  cggcctgcgc  cgttccgagc  agaatcgtct  4020
tctccggtcc  gccgcccgtg  ccccgcacgc  tgcggagctc  gagcacgcgc  acgggacgcg  4080
cgattcccgg  cctaggcgcg  gcgtctggca  gcatggcccc  cccggccgac  gagcgcgatg  4140
gcccgggcga  tatcgtacgt  gccggccacg  acgagcgact  gtgcgatctc  gacgagtccc  4200
```

```
actgcgccgc gccggcgccg gatcatcgcg gctgcgcgca agccggcgaa accggcaagc 4260
agcaggagcg cggtaccgac aatccaccac ccgccgcgct gccacagcgc gaggcccagc 4320
ggaatcgaga cgagcgccag cagatcgagg accgggatga tcacgctggg cagatcccgc 4380
atggcgagcc ggtgacggaa gctcacgcgc aggttgtcgc ggccgcgcca cagctcgccg 4440
aaaaacacgg cgcgaagcga tcggggatcg ccctggtgca cgctgcggag ccgcgggtcc 4500
gacaggatgc ggccgccgcc ccggcgcagc cgctggcaga tatccacgtc ttcacaggtc 4560
tcgaggctgg tatcgaagcc accggcgcgg tcgaaggcct cgcgcttcac ggcaagattg 4620
ccgctcggga gccagtccac gtcttcggtg ctctgcccac gacggcgaaa cgagtcgtac 4680
gcgcgctgca cccagttgcc gtcaccggga gcttcgtact gcgcgccggc agcgacgacc 4740
cccggggtct ggagcagatc gacggcgacc accggccagc gcgggtcgat ctcatgatcg 4800
gcgtcaacga acgccagcac gtcgccctcc gccgaccggg ccgcgcggtt gcgaagcgtc 4860
gccacgttca cgtcgggcag cgacagcact cgcgcggatc cgcgctctgc cacggcacgc 4920
gagtcgtcgg tcgacccgtt gtccgccacg acgatctcgt accagccagc cggcgcctcg 4980
gtgcgatgaa tcgacgccag gcaacggctg aggtgcgccg cgccgttctt gacggggatg 5040
acgaacgaca cgcgcggcgc cacggcaatt ttcgcggtca tcgatttcgt agcgcggccg 5100
gctttcgtag cgcggcccgg gttgggggcc gcgagaactc gaacgcgccc agatccggcg 5160
cccctgctct cggccgccga tcgaaatcga gcggcgcctg cggaatctcg atgccggcgt 5220
cgatcgcctc cctggcggct gccgacaggt gcagatcgca cgaggccgcg tttgcgaacc 5280
actctgcggc ggcgttcgtg acgttgttcg tcgcctccgc ctcaccgccg ttacggcgca 5340
ggatctgccg gctcgtcagg ttgttgcgga tgagcccgtt cgtcacggga aaccggacgc 5400
cgatcgtcca cggcgtcgcc gtgccgagcg tgaagacgct gttgtgctcg atccggaagt 5460
cggtggacgc gttggcttcg atcgcttcgt cggcccattc gttcaggttg cagatcacgt 5520
tgttccgcac gatgccgcgc tggtgatcga tctgcttctc gccgtcgcgt gcgtaggccg 5580
tgatccccgc ggccaggccg atggcgatgc ccctgaacga atcgacgatg acgttccgtt 5640
cgatgagcgt gtcctgcgag ttcgcccagg cgaggatcgc cgggccggat cgccagccgc 5700
ccgattgcgg gccgcgaatc cgcaggaaga cgttgtcgcg gatgacccac ccccgggccg 5760
ccaggatgtc cacgccgtcg gtgtaatcgg acggcgcgct cgtcgtgtac tcgaagcgcg 5820
agcacgccac gaggccatcg tccgcgaagc gtccgtcagt ccgacgctg cctttgagca 5880
attgctgtcc cgcatcgatg agctggacgt tgtgcaccgt ggctcgggac gcgccaagct 5940
cgccgcgcac ctgaatcgcg tggtacccga cgtggccgac cgtcagatcc gcaatcgtca 6000
cgtcagtggc gccaaccgag agcgccacgc cgatcgcgtc gccggtcatg ccgccgccgc 6060
gaatgacgac gcggtcccga tcgccggagc ggctccggag cacggtgccg gggacgttga 6120
tgctggccat ccggtcgagc cggtactcgc cgttctcgag cagaatcgtg gtgcgcggct 6180
tgacgcgttc gagcgcctgc agcagctcgg acgtccgccg cacggtgacg acctgtttcg 6240
tcgccggcgc cgcccatccg cagaagtgcg gcaccggcgc ggcggcctgg ccggacggca 6300
gaagcgccac gagtgcgccg gccacgatga ggcgagtcat gccgccaccg tgtcgacgtc 6360
gtgttcgcgc aaccagagct cgaggatgag caggatttga aggaggtagc cgcgcgactg 6420
tctcccgtga cgcgtgtcgt cgatgatggc gcgcaacgtc tcgggcgca ggaacatccg 6480
cgtgcgcgcc tccggggcga acagctcgcg ctcgacggct tcgagcagtc gcgtctgcag 6540
ccagcgatcg aagtcgtgat agtgacggta cccaggcgcg ttcaggcgct tcagcagggt 6600
gttgaccttg tcgaacacct tctccgccag cggaccggcg tcgccgcgcg cgccggtgtt 6660
cgagttccgc accttgagca gcgccgggtt gccggcctgc gtgatcgcgc gatgaagcgc 6720
cgtgtcgtcg cgccatgccg gatcgccgct cagcaggacc tgcaggaact cgcgatccat 6780
gaacggcagc cgcacgtcca cgacggtgcg gaagagatcg agcgagggaa tcgtgaagcg 6840
ccgatgatgc tcacgaagat agagatagct gctgagctcc gcgggtgaga gctccacggt 6900
gtcgaggagc gcgtgcagcg agtcggccgc gccccgtccg gagaggcgcg cagcgtctgc 6960
tgtgaagagc gtgtccagcg gcaggccgct gctgatgtag ttggcgcgcg acgtcatgta 7020
gtcgatgaac tcgccgcggc ccgtcatcgc gaacacgtgg ccgtcggtgt ggagcggcca 7080
```

```
cgccaggctc gtcttggcga gctctccgcc gtggccgcgc agcaggacct cgataccgga 7140
gtcccgcagg aaatcgatcg ccagcatctc ggtcagtccg tggctcagat acatgccatc 7200
ggtgagcgag accatgcgcg cctggttcgg cagaaagtcc ttgaggtacc ggtcgtcgag 7260
ctcgaagaag cggtggtcgg taccggtgag cttcgacagc tgttgcgcga tcacctgatc 7320
ggcgcagccg gcgacgccca gcgtgtacgt gatgagcccg ccaacaccgt tcaccgcact 7380
cagcacggcg cggctgtcga ggccgccgga gagcgacagc ccgaagcgat gcctgcccga 7440
gagcgacttc tcgaccgcgc gcctgaacgt cgtgtgcacc gcgtggacgt acgcgtcctt 7500
cgtcgtacgc ggtccgcggt agagatcggc cggctcgaaa taccgggaca cgctcgcggc 7560
gcctgacggc cagtcgtacg cgagcacggc gccgggggtcg aggagcgtga cgcccgccgc 7620
cagggtcttg tcgcccagga cgaagcccag cgtgacgtaa tcggcggccg cccggaggct 7680
gagagccggt ttgcgcgccg aggtcctgag cacggaagcc agatccgatc cgaacgacag 7740
tccggccgcg tcgacccgcc agtagatcgg atacgagccg aacggatccg ttgcaagcag 7800
cactcgctct cgcgccggat cgatcacggc gagcgaaaac gcgccgtcga gccgggcggc 7860
catggcaggc cccagccgtt cgtacaccaa agccaggaga tcggcgcgcg tccgttcgc 7920
cgaaccaaac tcgctccgga gcgccgcttc gttgtagaga tcgccgcgaa agatcaccag 7980
cggccggctg acgtcgcgcc cggccgccac ctcgtcgagg aagccgagcg tgccgagccc 8040
gaccgcccat cggtccgacg gatctcgata gacgaccgtg cggcgcggcc tgcccgactt 8100
cgccgctgac gcggcgtcgg cgcggctgaa ggcggtagat gccggctcat cgcggcaaac 8160
gacgccgagg aggccgttca tcagacgacc ccgcgggcag acacgcgcga cggaacaaga 8220
ggtgtcgtct gagcggccgg aaccggccgc gcggcgtgcg ccgccaccat ccggttgagc 8280
gcaatcgaca agccgaggaa gtgccagagg atctcggtgt actgaaagat gacgaacgtt 8340
ccgcccacgc agaatgcaat gagggcggtc tggacggcca cgcgtacgg cgccagctcg 8400
gccagggccg gatcccgctt cgagacccgg cggaccctgc ggcaggcgat gagcgccagg 8460
aggagagtga gcaggaacgc cgcgaggccc acgaacccaa gctcagccac gagaccgaac 8520
cacgcgctgt gaaccgagcg ccccttgccc caggcgccgt ccgacgtgtc gtagtcgttg 8580
taggcggccg tgaaggcatt gtgaccgacg ccgtaaatcg gccggtcatc ggccatgcgg 8640
agcgcgaccc tccagaagtg cagccggctc tgggccgatc gctgatcgcc aaccgccgtg 8700
tcgctcgcgt cgatgctcga gtcggacgag ctgatcgagc tcatgcgatc ccagaactct 8760
tccgtcatca cgagcgacat cgcgacgaag accgccgccc cggccacgag gacgagcatc 8820
ttgcgcttcg agtacaggaa atagacgagg ccaagcacgg cgagcgtgag gaagccgccc 8880
cgggagtacg tgctgatgcc gcgatagagc accccgatga gcagcacgcc cagcgcccac 8940
ttggcccatt tcctcgcttc cgtctgcatc agcgtgacga tgagcggcac gagcatgaac 9000
atgccgacgg cgacgccgtt gttgtcaccg agcgagtaga cgtcgttgaa gttcttgatg 9060
ccgggaatga acaggagctg accccagccc tgcttggcgg cctcgaaccc gagggagagc 9120
gagatcacga ggaagacgag ccgcaaccgc ttgacgtccg tcgtgaggac tgcgagcagg 9180
tacgtcatca cggtggactt catgaagtcg atcatgtagc cccacgcgaa gtcctggtag 9240
ggcgacgcca gcgtcgacga caggctgtgg agcccgaaga aggcgaggag cgccagccgc 9300
gcgtcgatcc gcagcgcctg cccggacagg aacgcgatcg cgagcgtgta catgcccgcc 9360
aggagcgaga tgttcatgtg gatgagaaag tcgctccaca cccagagctc cggccggaag 9420
tacgcgatga acaggtagaa cagaatcgcg tagaacggcc cgcgcagcgc gtggaaggcg 9480
ccgaaggcga ggagggcgag gacgaaagcg gctcgaagca ttgccgtgtg ttacctgtta 9540
tcggtcccga tggtccagag cggctgcgtc cgggtctcct ctacattgaa caccttgtag 9600
agggccggaa tcgaggtgaa catgagcacc acgaaagaa tcgccgacac caccatgtag 9660
gcgaagaacc cccgctgctt gtagagcttc tccggattct gcacggggct gttcggcagc 9720
atcccgatgt gcaggtagta ggcgaacatg cccgccgcca ccggcgcgaa caggatcagc 9780
tcgaggtgat agcggacgat gaagatgccg ccgaagaggg cgcagcaggt ggcgtagaac 9840
accatgctca cgagcagccg ctcctcgttg tagaaggcaa acgacttccg gtacgaggcc 9900
gcgaccgacg cgttgccgat gtgccggaac tccgcgtacc gcttcgtcgc catgaagaac 9960
```

```
gcgcccacca tccagtacga gatcacgagc gacatcggcg gcacgcgatc ctgaatcagc 10020
gggaaccagc cgaggaggag ccggacggcg ttgttcgccg actcgctcag cacgtcgagg 10080
tacggccact ccttcgtgcg aatgggcggc acgttgtacg tgacgccgag cacccagagc 10140
gccagggccg acagcgcgaa gtatctgttc acgagcagcg cgagcacgaa tccggccacg 10200
cccacgagga tccactcggt gtagcccgcc gccggcttga tcttcccgga gggcaccggc 10260
cggtgccgtt tctccggatg cagcagatct ctcgggccgt ccaggagctc gttcagcacg 10320
tagttgctcg aggcgatgag gcaggtcgcg gcgagcgcga gcgcgagcgg cgggacggcg 10380
gtccagccga agagctgcgg ttcgtagaag aacgccagca gcacgccgag cagcatgaac 10440
gcgttcttga accagtgatc gatccgcgcg atctggacgt acggccaaat ccgcgaaccg 10500
gagctagcgg ctgacgacat cgttgagctc ctcgagccgc cggatcgcga accatccgcg 10560
gccggctgtg acgccgtggc ggccggcgat gagggcgcag gagagtcctg cggccatcgc 10620
tccttctccg tccacgtccg gtcgatcccc gacatacagg acctcccgag gtccgagtga 10680
ccacatctca cacgccacat ggaagccgcg cgggtgcggc ttcagggcgt tcacgttgct 10740
ggcggtcgtg cagagcgcca gcgaaaaatg ctgcgacgtg ccgagcgcgg cgagcttgct 10800
ctcgggcgcg tagtccgaca gcacgccgag cctgagcccc gccgcgcgaa gcgagtcgag 10860
cgcgccgagc aaccccgggc gccggcacca ccgcagatac ttgagcggcc ggcgcaccat 10920
ccattcgttc accgtcgccg ccacgacctc gttgtccagc ccgagcgcct gcgcggtgcg 10980
cgcgatctgg cgcgccgcga gcggctcgtc ggccgcgccc aggcgccgga gatcctcatg 11040
cgcccggcgg tactcccgca cgatgcgcgc cgtgtccact ccgcgccgcc accggaacgt 11100
cacgagcggc gtggcggcga gctccgctgc catggcggcg cgcaacggac cctgccggta 11160
cagcgtgccg tccacgtcga acagcacggc cttgaccccg cgcaacggct tcactctgcg 11220
aggcccagca tccgcagaat cgccgacacc tttgacgcgt cgcgaaggaa cacgtcacga 11280
acccacacgg cagccagccc cgccaacagc accgccgctg ccgcgcccgc tgccgtccgc 11340
cagctcccgc gccaggcgcg cgtgtcgatc gcggcggcgc cgtagcagag cagaatcggg 11400
acgagcggca ggtgataccg gggatgtccg aagacgatcg cgtgaatgcc gctcatgaac 11460
acgatcaggc tcaggagcag caggtgcgcg cgccagttga ccggccgcgc gaacgagatg 11520
ccgatgacgc cgagcaccat gaccgccgcg tacgatccca tgatggacac cgccccggcg 11580
aagaacagcc accgcggcgg acggtagaac ccatttccga cgccggcgat gaagtcgcgc 11640
tccagccccc agaagtctgc gaacttcagc accgaacggc gaagcgtcgt tcccggatgg 11700
gccttcatgt agtcgagcgc ctggcgctgc gcccacttct ctttcgtgcc ctccgtccag 11760
cggctgccgt ccggcgcgcg aggcggcatc gtgcgcgacc attccttgtc gcccgtgagc 11820
gagatcgcgt cccacatgcg atcctcgggc gtgtacgcgt agttgcccat catcaggttc 11880
agcccgccga gcgtgtccac daccgtgaac gagcgttgca ggagcgtgtt gcgcacgctc 11940
cacgggccga cgatcgcggc gtatccggcg aagagaagca cggccacgcg cagacgcgcg 12000
cgaagcggga tgcccagacc gacgagcgcc gcggccatga ggatcgcgac gtacggccac 12060
atgatcgagc gaacgagcgc ggacgcgccg agcgcgacgc ccgtcgcgag cgcgatccac 12120
gccgagtgcc gccctgatgg gcgatcgatg agcgtcagac atcccaagag caacaaaga 12180
aggagtgtga tgaagagcgt ctcggacagc accagcacgc cggagaacag cagcgaggga 12240
tagaacgcga acgccgccgc ggcaatgagg ccggcgcgtg cgctgaacaa cctgcgcccg 12300
atgagataca cgagccagac accgagcagg ctcaacggaa cctgcgccag ccggaccgcc 12360
gtcagcgatt ccgagcccgt gaccgcccag acgcccgcga cgaacgcggg aaagagcggc 12420
gcccggatgg acgtcggttc ccccggcccc cacgcgaacc cgttcccggc gacgacgttt 12480
ctggcaagcg tcgcgtagtg ctgctcgtcg cgaatctcga gcgcgacgtc cttgaacccc 12540
caaatcaggc cgagccggag gaccagggcc agcgtgagga tgacgagaat tgcccggcgc 12600
tcccaatggg tgcgctcggc ctcgtcgcgc cgggggcga cgattggcca gggggccgac 12660
gatgggagtg ggggagcctg catcattcgc taaaacaagc agttacggtc aaacctgtgc 12720
caccggggaa tcagcaaaga ctctgccaac gggtccgccg gcggcgacgc ccggcgcgcg 12780
ttttccggcc ggaagtcggc ccccaaggcg accggctggc gcacagccca gtagaaatct 12840
```

```
cattctggga tgacttttgt ccatgatcca tgaccatttc gaggggcctt cgtgacgccc 12900
tggggatggc ataagcttgg cacctgcgat ggcgacgtgt cgatgatcag cggccgtgag 12960
gagctccaga aggcctatcg ggacgaccgg gtcgcccgtg aatacgtggc gcggcggttc 13020
cagtcaccgc tggggcgct cctccattcg cgtcagatcg gggtcgtgcg agagctcgtc 13080
cgcgcgcagg gcatccggcg cgcggcggag atcgcacccg gtccagcgcg gctgaccgtg 13140
gacatcgcgc cgctcctcga ccgcgtgacg ctcgtcgacg cgagcgccca gatgcttcac 13200
gaagcgcgcg agcggctgcg cgaccggacg ctcgccgcgc gggccaatct cgtgcaggcc 13260
gacgcgtttc gtctgccgct cagcgggccg ttcgatctcg tctacacctt caggctcgtg 13320
cgccatttcg agcgcgccga tcggctgcga ctctaccgcc agattgcgac gattctcgga 13380
cccggcggct ggctcgtctt cgatgcggtg aatgagcgttg tgtctgcgcc gcttcgcgcc 13440
aacgcgaagc ccggcgagta cgagcactac gatgcgctgc tcacaccggc cgggctacgc 13500
gacgagctgg ccgaggccgg gttcacgatg gtgtcgctca ccggtgcgca tcgccggttc 13560
cccacgctca tgaagtgcca gatgtatctc gcgccgaggt caacggtgct cgcgcgcgcg 13620
gcgatggagg tgatcgatcg gctcggcggc gaaccgctcg agtggatcgt cgtatgccgc 13680
cgcgggtgac ctactggacg ggtacctggg atccgatcaa agaagcgatt tcgaaggaag 13740
tgaacgccct ccgcaccggg tcgcgggcga acgcgacggt cgtgtcgttc tcgccggccc 13800
agcgcagccg tctcgtgccg aaagatcgcg tgctcgtgtt gtccggccgc gcctggatcg 13860
cccttcgggc cgcggccgcc gtgctcgagc cgcgcggcga cgtgacgcac gtcttcggcg 13920
gcgggttctc ctggcatctc ctgagggcgc tggggcggcg gccggttctg ctcaccgccg 13980
tcaaggaaga tggcgctgcc gatctgcccg gggcgatcgc gcacgtggcc gtcgaagtcg 14040
gccgcgctcg cgacacgtgg atcgcggccg gtatgtctcc cggccgcgtg cacctcattc 14100
atcccggcgt ggacctcgac tggtatcggc cggtgccggt ctcgtcggag agattcacgc 14160
tgctcttcgc cagcacgccg tccgatccgg ctgagatcga gccgcgcggc attccgctcg 14220
tcgtcgagct ggcgcggctc cgtccggaga tcgacgtgct ggtgccgtgg cgccggtggg 14280
gcgacgtcga cgcctccagg cgcgcgctcg acgccctccg tccgccaccg aacttcatcg 14340
tctcgtacga ggacgcaccc gacatgcgcg cctgttacgc gcgggcgcac gccaccgtcg 14400
tgtgcttcgc gggcggcacc ggcaaggcct gtccgaattt cgtgctcgag ggattggcca 14460
tgggacggcc gtgtctctcg acgcgcgaga acggcatcgc ggagatcctc gatcgatcgg 14520
gcgccggcat catcgccgcg cgcgatgccg cgtcgctcgc cgcgagcctg gacacgctcc 14580
gcgccggctg ggccggctgc gccgagcgcg cgcgccggct cgcagaggag acgttcgacc 14640
tcggccgctt ccgcgccgcg tacgacgcgc tgtacgtgga ggtggcatcc gccgccgctg 14700
gcgagcacgc gaggcgcgcc tcttgacgaa gggactcgcc ggacggctgc tcgggtgggc 14760
tcagaaggcg aacgatttcg ccgactacgt ccgccatcga tcgcagatga agcgcgacgc 14820
gcgggcgaat cgccgcaggc tccatcgggc gctcggcggc cgcgtgaacc ggatcctcat 14880
cgtgcggctg ccagcatcg cgacgtcgc gcgcgcgacg ggcgtcctcg gggccctgcg 14940
cgcgcagtac ccgcaggcga cgatcgactt cctgacgtcg gacgcgaccg tggccgtcat 15000
ccggcaccac cccgcgatca acaccgtcta cacgctccag gatctgtcgc gcctgtccga 15060
gtacgactgg atcatcaacc tccagaacct cgtgccgccg gcgagctttc tcagaacgtc 15120
cgggctctcg tacccgcaga ttctcgagca gctctcgagc cgcagccgtt ccaggatcat 15180
caccggccgt cacctcgaga acggctcgga gtccacgccg acgaacatcc tctactgcat 15240
ctgtgagatc gaggagcatt tcctcacggc gctcgtcccg tacgacacgg cccgctaccc 15300
ggtgaccgcg attcacgtgg acgcacaggc acgagcggcg gcacgatcga agttcgcgtt 15360
cccggagaac cgccaggtgc tggcgatctt cctcgggtcc aactcggtcg ggtgcggcgc 15420
cgatgagggc tttcgcacgt actccatcga ctacgtgaag cgcctcgtgc gtcacttcga 15480
agatcggttt gcgatcgcca tcatcgggca gtcccaggtg cgcaacccgg ccgaacgcgc 15540
acagtaccgc gagatgctcg gccgccatcc ggacgtgatc gatctcgtgg acaagacgtc 15600
gctcgacgag ctcgtggccc tcatggactc gttttccctg ctcatcagtt gcgactcgag 15660
ccccgtccat ctggcgatgg cccgcggagt gccggtcatc gggctgtacg tcaacgacgg 15720
```

```
gacgttcagg atgcacccga cgctcggcga cgagcgcttc gtggccatca acagcacgcc 15780
gccgtgcttc acctactcgt ggcgatggaa gttcttctgc tcgacgtgcc gcgacccggc 15840
gaccagatcg cattactgtc acaacgaggc gttcgtgttt ggcgtggacc gcattccgct 15900
ggccgcaatc gaccgcgcgg cggcgcggtt gctcaagggg agctccgatg caggacctca 15960
ttcgcgcgca gtttgaagag tcgatcgaga cgaagcgcaa agcgctctcg ctcgtcgatg 16020
cggtggccga gtcggcgcgc ctcctcatcg aggcgctcga ggccggacac aagatcctcg 16080
tgtgcggcaa cggcggctcg gccgccgact cgcagcattt cgccgccgag ctcgtgggcc 16140
gcttcgagaa ggagcgccgc gcgttgcctg ccgtggcgct cacgaccgac acgtcgatcc 16200
tcaccgcgtg gccgaacgac tacgcgtacg agaccgtgtc agccggcagg tgagtgcctc 16260
ggagcgccgg gcgacgtgct catcggcatc acgacgtccg gcaactcgcc caacatcctc 16320
gccgccatgg acgaggcgga aaagaaacgc atgaaacgaa tcgcgctcac cggacgcggc 16380
ggcggcaaga tgaagagtct cgccggcgtc tgccaggtca tcgtcccgtc agatcagacg 16440
tcgaggatcc aggaagtaca cattaccgtc atccacatct gggcgaagtt gatcgagaaa 16500
gcgctcagct gagctggcgg tagacgtcga acgtgcgatc ttcgaacgtc gaccaggagt 16560
gctcgcgggc gatgaggtcg gcggcggcgt cggtgaggcg gcggcggagg tgctgatcct 16620
cgaacaacag ttgcagtcga ttcgcgtagc cctcggcgtc gtccgcctgg acgagaaagc 16680
cggtcttccc gtcttccacg acttcacgca cgcccgggat ggtggagccg atcacgggca 16740
cgcgggccga ctgggcttcg agcaccgaga gcggcagccc ctcgtgcgtt gacggcagca 16800
ggaaggcgtc ggcggcctgg agcagcaccg gcgcgtcctg gcggaagccg agaaagcgca 16860
cgcgatcctc gagcgcgagc tcggcagaga ggcgcctcag cgcctgctcg tgttccttgc 16920
cctcggtccg gtcgctcccg accagccagc actcgaccgg caggccgcgc gacatcagca 16980
gcctggtcgc ccggattgcc gtagcctgcc ctttgtgtgg cgcaagattc gcgagcatga 17040
ggacgacgaa gcggtcggtc tcgaggccgg cggccgtacg cgcggcttgc cggtcgccgg 17100
ccacgaactt ctcccgatcg accgcgttcg gaatcgcggt gaccggcacc ttgtcccggc 17160
gtccctgctg cgccgcgtcg atcatcacgg cgatgtagcg cgcacacgtg acgatgtgct 17220
gcggcgggct cttcaggacc caggcgatct cctcgggatc cggttcgatg tgaaagtgca 17280
cggccacgcg cgcgcgcgcg gcctgcagcg cgggtctcat gaatccgtac acgacaggat 17340
tgtgcacgtg daccaggggc cgcccgacga gacccaatcc cggaagcatg cgcgcacacg 17400
cggcgagctg ttgcaggccg ccggcccgca gcgcgttcag gtcatacatc cgccgccgga 17460
cgccctccgc gtcgagcgcg gcgcccgcag gaccctcgcc tgtaatccat gccgtgcacg 17520
ggacgccgcg gcgtttggcg gcggccgaca accggatcgc cagcaccgcc gcgccgccga 17580
ccagatgtga gacgaggatc tcgtgcaggc ggggggactg attcatggtg gcgcgcgatg 17640
cggccgggcc gccagcctac gttactcgcg tcgcaggtcc gcccgcgcgg tagagcgagc 17700
gcagcaggct gcgaagtccc gacgggagga agcacacggc gtaccgcgtg agatcggtgc 17760
gccgcatcgg caggcccgcc tgaccgaacc gtttccgcgc cgcgtgaaac tcgccgcggt 17820
cgtactcgag gtggccgagc gcgagcagga gacgatggcg atgatcgctg gcggcgcgat 17880
gcgccgactg atcgctcccc cgcgccagga accgctccag cgtcctgacc atgccttcga 17940
gcatcacgat ggacgacttc gacatgttgg tgtcgtgcag cctgtaatgc gtcagggtt 18000
ccgggatgaa accgaccggg cccttcgccg agaggcgcag ccagaggtcc cagtcctcga 18060
cgccgtggat gccggtcgcg aacatctcct gccggagcac ggcggctttg accatgaccg 18120
acgaagcggt cacggcgttt cgcaggacga ggtcggggag acagcgcccc tcgagcggcg 18180
gaagatcgag cgtgcgcagg acctcgccgg cctcgtcgat gacggcgcgg cccgtatgaa 18240
tgaagctcca gtcggcgtgc tgtcggaaga gatcgacctg ccggcgaatc ttggccggat 18300
cccacgagtc gtcggcgtcg agcagcgcca cgaattcccc tcgcgccgcg gcgatggcgc 18360
gattgcgggc gccggcaagg ccccggttgg cctgagagat caccgtgatg gcgccgccga 18420
atctcgccag gatgtctgcc gtgccgtcgg tcgatccgtc gtcgacgacg atcacttcga 18480
cgggctgata ggtctggag agcgcggact gcacggccct ggcgatgtag cgctcgccat 18540
tgaacaccgg gacgatcacc gacacgaggc cggggatatc ggtctgggcg gccggcgagg 18600
```

```
gctctcgagc ggcttcagcg gccatggccc tctgcactgg cgtcgcgagc gaacagcatg 18660
acgtttctcg aaagccacgg agacaacggc gcgctcttca tccgctcccg cagcgtctgc 18720
gatgccgggc cgaggtagct gaacccgcgt gccgcgaaga tcttctgcca gtactcgagc 18780
ggctgttcgt tgatgtggcc cattccgccc tggcccggct gcgcggcgct gaagagcacc 18840
agcggcgccg tgttcgtgag cacggcgacg agcttctcgc caagcggcgg cggaaggtgt 18900
tcggcgacct cgaggctcat ggagaggtcc gttgcgtccg cggccgtgga caccgccgtg 18960
agatcgaatg agttcagcgc gacgccgagc ttttccgcgg cgtaccgtcg cgcgtactcg 19020
gatcgctcgt agccggccgc cttcacgcct cgtgcgagga acgcgtgcac gaagtgaccg 19080
gtgccgcagc cgaagtcgtt catcgatcgg atgccgggtc gcaacgccag gaccccatca 19140
acgagcacgg gcgccgcctc ggcggcgacg tgcccgaaca gatccatgac ggccgcgttg 19200
tacgtccatc gctcgaagcc cagccgttcg ccgatccggg acatcttcag tccggcctgg 19260
caacgccccg accagtagag ccgcttgagc tgggtcacga tcatgggttc gtcccgggcg 19320
cgacgcgaac gcgcggcggc gcatgcagcg acgccgagaa ctccacatcg gccacgccgg 19380
cccacgatcg atgttccttg accgccagcg tcgcaagctg ccgcgacggc acgaggaacg 19440
tctgggtcgg gttgtgaaac acgtgcaccg acaggtagta gtggccgcga acgagatgcg 19500
cgcggaaatg gaggtcgagc gagaactcgc ccatgaacgc ggacgcgtcc ccgagatccc 19560
gccggtggat gttgccgtcg tagacgacga gctgatcggt ggagcgttgc accatgaagc 19620
cgaacgtgac gtcgtccagc ggctcggtac ttctcgcggt gacctggagg ttcagctgct 19680
tccccggcgt cacttcgctg tgcgtgcccg acgactcgtc gccgacgaag agcgagccgg 19740
acacgatctc gacggggctc ccgacactcg gctgttccga tttcctggac gcgcggacgt 19800
acgcgtcgag gacttcgatg gccgtgccgg cggccttcac cgagccggcc agatacaccg 19860
catgctcgca gagggtcgcg accgcctgca ggttgtgcga gacgaggacg atcgcgatgc 19920
cctggcgctt gaactcctgc atgcgatcga agcacttctg ctggaacgtc atgtcgccga 19980
cgctcaacac ctcgtcgatg atcagcacgt cgggccgcag gtgcgcggcg atcgagaagc 20040
cgagccgcgc gttcatgccc gacgagtagc gcttgacggg cgtgtccacg aagtcggtga 20100
cgccggcaaa ctcgacgatc tcgtcgaacc gctcggcgat ctcgacccgt ttcatcccca 20160
tgatcgcacc ctggaggaac acgttctcgc gcccggtgag gtcgggatgg aaccccgcgg 20220
cgacttcgat gagcgcgccg gcgcgccccc ggacgatcga gcggccggtc gtcggcttca 20280
ggatcttcgt cagcaccttg agggtggtcg acttgcccgc ccgtttgcg ccgatgatgc 20340
cgagcgcctc gccgcggtgg acttcgaacg agacgtttct gaccgcccag aactcctgtt 20400
cgtccagcgc gtcggcggga cgcggttgaa acagcccgcg cacgagcgcc ggcacgaggt 20460
cccgcaggct gtcatgacgc tcgccccgcc ggaacttctt ggagacgttg tcgaagacga 20520
cgggggccgt cacgtcagac gttctccgcg aattcgaact cggagcgatg gaagaacagc 20580
cagccgaaga cgagcgtcac gaccgagacg atcgccgtga cgcccagcat cacggacggc 20640
ggctggttca ggagcagcac ggcgcggaag ccgtcgatga tcggcgtcat cgggttgagc 20700
tgcagggccc actggagctt gccgcccacc atttcgaccg ggtagaccgt cgacgacgcg 20760
aacatccaga cggcgatgac gacctcgaag aggtatttca cgtcgcggta gaagaggttc 20820
gccatcgaca ggatcagggc gatggccgtg gtgaatatca cgagcacggc gagcacggcc 20880
ggcagccaga ggatgttcgg tccgaccggc acgtggtagt agatcatgag cgcgacgagc 20940
accacggagc cgatcgcgaa gtccacggcg ctcacgatca cggcggagaa cgggaagatc 21000
tccctgggaa agtacacctt cgagacgaga ttcgtgttcc cggtcaggga ggtcacggag 21060
aagcggagcg ccgaggagaa gaagttccag acccacaggc cgcagaacgc gaacaccgga 21120
tacgcgaccg gcgtgtcgat ggccgcgacg cgcatgaaga tcacggagaa gacagccgtg 21180
ttgatgagcg gcatgaacac ggcccacccg aagcccatca ccgactgctt gtagcgcagc 21240
agcaggtcgc gccgcgtcat ctgcatcagc agctcgcgat actcgaactg ctcggcgagc 21300
atctcacggc aatcggccag gaaggcgcgc atcgtcagcg gtgtcagtcg cggcgtcgtg 21360
ccgggcggat catgagggca ggtagcgctc gagctgcccg aagagcaact gcacgaactc 21420
ggtcgcgcgc cgctcggcgg cgacgtcgcc gtttttcgctc gacggcacgg ggctcgtcac 21480
```

```
gcggacgagc gcggcgtccg atcggttcag ccgcatcgcg tcgtacacca tgtagatctt 21540
gctccagtac tcgctggcga tcacgcggcc gtggctctgg taccagtaca gcacgatttg 21600
ccgctcgatc cccttctgga tgacgtatcg gttgatctcg atgggctcgg tgcggcccgg 21660
caccttgacc agcgcgcggc cggtggatac gggctcccag ccggcgcccg gcaggcagtt 21720
catcggcgag tggatcgtgt cgccttcgcg ctggctctgg tagtagccga tgtagagcga 21780
cacgtagggc tcgcgcggcg ccaggtagat ccggttgatg tactcgtcca cgccgagcac 21840
cgccatgacg tccttggtga agggcgcgct gtcgcgtccg gtccaccgct ccatctggaa 21900
cgggacgttc gcgagcggct ggcgaagcgg gacctgctcg gcgcgcatcg agcgcgcggc 21960
aaacgccgcc gtgccgagga aacacacgct gagaatgatc agtcgtctgg tcatgaatct 22020
ggactcacac cgtggccggc gcggtcaccg ctcggggcgg aggcgtgaag cgccgcagga 22080
cccagaccac gacgaacaac atgacgaacg cggcaagaaa gacgagccag cccgagaacg 22140
tgtggaagaa cccctgcgcc gcggcctcgc cgtagtagtg cgccgccacg cccgtgccgg 22200
ccacacgcgc gccgttcgcg ataatcgcga tcgggatcgt cgagagcgca atcgccaggc 22260
gcatgccgac gcgcggatcc gtgaagtagc cgtacacgat gcccagcgtg agcagcgaga 22320
tgagcgaccg gatcccgctg cacgcctcgg ccacttcgag cgtcgtgtgc gcgagcacga 22380
tcacgttgcc ttcgcgcagc accgggatgc tgagcgcgga cagcacgccc tcgcccacgc 22440
gagacgccag gagctgcagc ggaaatgcga tctgattgaa gatgatcgac gggatcggga 22500
tcatcagcac caggaacgcg agcggaaacg cgagcaccca gagcttgcgc cagccgagca 22560
ggaacacgat cgcgccggta agcgaggcta gtatcgacac gcgcgtcaga aagagctcgg 22620
cgccgagcaa ccccgcgacg agcagcgcga gcccggcgag cacgatgacg agaccgagcg 22680
cgctcggcgt gtccggcagc gccgagagac gcgcccggcg ctcccaggca aagtaggccg 22740
cgagcggtac gatcagaatc ccgtgggagt agttgtcgtc gttgatccag tcctgcgcca 22800
accgcacgat cacctgccca taaaggagca ggaaactcac ggcggcgata ccgagcggaa 22860
ccagtctggc ctgaagcaca ctcatagcga tcgggacgaa tatagcaaag ggctctcccg 22920
ggcagaactt gggcgccgaa acacgtaacc atcacgctca caggaagtta gcggctaaat 22980
gacgggattg gccggctgga ggggcctcgt gaatctaccc aatggatgac aaatatccct 23040
atcaactcct ccttttgact gcgtagcacg gaagatcatc ggcctcttca gatgcaggac 23100
aaatgtcatc tgtcaggaag agtaatgtct ctggccacag cgccggtcaa gtgaaatccc 23160
ccgacagaat cgcgtaccac cgcgactggc atccggcgag aaaagtcgcc tcgtcggccc 23220
gaaaactgca ttacaggaac gccaggtcgg cagccggttc gttttccgtc agtattttg 23280
ggcattgcag gcgctttcgt ctgccggggg gcggaaccgt gcgaccgggc cgatcacgcg 23340
tggagcaggg cagggacagg catggatggt tccgtgagac agccgcgttc gggagacggc 23400
cgggctcggc gggagctcgt ccacctggcc gtggaccgcg cggcgcgcca atttcctgac 23460
gcgccggccg taagttttga cggcgtctcg ctctcgtatc gcggcctgac tcgccgcgcc 23520
aacgcgctcg cgcaccgcct gcagcgtcat ggcgtcgatc ctgacgtggt ggtcggcatc 23580
tccctcgacc ggtccatcga gctcatcgtc gggatcctcg ccgtgctcaa ggccggcggc 23640
gcctacgcgc cgctcgatcc tgcgtatccg aaggaccgcc tcgagtacat ggtcagggac 23700
gccggggcgc ggatcgtgct cacctcgcga cgcgtgagcg gtctgatccg cgtcgatggg 23760
gtcgagctca tctgcctcga tgcggacgag cccggggcgt tggcggaaga gggccatgcg 23820
ccggacgccg gcgtcgatct cgaccacctc gcgtacgtca tctacacctc gggatcgacc 23880
gggcgcccga agggcgtcgc gatgccgcat cgtccgttgg cgaatctcat cggctggcag 23940
gtcgacgcat cggccgtcgg cgccggcgcg cgcacgctcc agttcacttc gccgagcttc 24000
gacgtctcgt tccaggaaat ctttgcgacg ctcgcgaccg gcggcgagct cgtgctcgtg 24060
tccgaggaca cgcgccggga tccgcgcgcg ctgctgcggg tgctgcgcga tcgatccgtc 24120
gcccggctgt atctcccgtt cgtcgcgttg cagcagcttg ccgtcaccgg caccgacgat 24180
ccggatctgc cggcgttgcg tgaggtcatc acggccggcg agcagctccg ctcgacgccc 24240
gcgctcgtcc agtggttcgc gcggcatccg cagtgcacgc tccacaacca ctacggcccg 24300
gccgaagcgc atgtcgtcac tgcgcacacg ctgactagcc cccccgcgga atggccggcc 24360
```

```
ctgccgccga ttggcgtgcc gctgcccggc gtcgagattc tgatcgtcga cgagcatcgc 24420
cagcccgtcg cgcacggcaa cgtgggcgag atcctcattg gcggcgtgtg cctcgctcgc 24480
ggctacttga atcagccgca gcgcacggcc gaacggtttc tctcgcatcc gcttcgcccc 24540
gatgctcgcg tgtatcgcac cggcgatctc ggccgtgtgc gcgacgatgg cgcgatcgag 24600
ttcgtcgggc ggacggacga ccaggtgaag attcgaggct tccgtgtgga gccgggcgaa 24660
atcgagaccg tcctccttca gcatcccggt gttcgggaga ccgccgtcgt cgcgcacgac 24720
gacgcgtcgg gcagcaagag gctggtcgcg tatgtgattc cgtcagccga tcaggcggcc 24780
accggagccg aggcttcggc acaggtctcc ggatggcaga ccgtgtggga ggacacctac 24840
cgcgcggcgg cggactcgac ggatccggcg cttgccgcga ccggctggcg gagcagctac 24900
acggggatgc cctacaccgc ggcggagatg cgggagtggg ccgatgagac cgtcgcgcgc 24960
atccggtcgc atgcgccacg cgaggtgctc gagatcggct gcggcacggg catggtgctc 25020
ttccgtctcg cgccccactg cgctgtctat agcgcgacgg atgtctctcg cgccgcgctc 25080
gaccacgtcc gggcccacgc gacggcctgc ggcgccagcc acgtcgagct cctctcgcgc 25140
gaggcggcgg actttaccgg cgttccggcg gccgcgttcg atgccgttgt gatgaactcg 25200
gtcgcgcagc acttgcctgg cgtggactat ctgacgcgcg tcctcgccgg cgccgcccgc 25260
accgtgcgtc cgggcggatt catcttcgtc ggcgacgtgc cgaacctgcg cctgctcgag 25320
atgtttcacg cgtcggtcga gctcacgaaa gcgccggcgg atcttccgat tgccctcgtg 25380
cgcgagcgcg tgcagcgtca gatggcgctc gagcgcgagc tcgtgatcga tccgatgttc 25440
ttcgacgccg tcatggaccg catccccggc atctcgggag cccaggtcca gatccggcag 25500
ggtgccttcg acacggagat gaaccggttc aggtacgacg tcgtgctcac ggtgggcgga 25560
cggccggccg ggcgccgtcc gtccgtcgcg ctcgactggc gtacggagcc ggtcgccgag 25620
cgtcagctcg ggcaactgct cagcgccgat gcgccggacg cgatcgagat caggaacgtg 25680
accaacgcgc ggctgcgtca gccggccgcg tgcttcgatc tcgtcaagcg tgccgactgg 25740
accggcacgg ccggcgagac gcttgagagg gccgccgcgg acacgcgcga cgcgatcgac 25800
ccggaatcgt ggcggcgtct cgcgaacgcc gccggctacg acgccgagtt cacctggtcg 25860
gccgagatcg gcgccatcgg cttcgatgcg cggttcgagc gccggggcac cgcgtcgctc 25920
gccgctcccg tgactcgtca cgcggttccc gccgcggagc cgcgggccat tcagcagtac 25980
gcgacgaatc cgctccgcga tgcggtggcg cggcgcctca cgcccgagct cggcgattc 26040
ctgcaggagc gcgtgccgga tcacctggtg ccgtcggcct tcgtcgtact cgatgcgatg 26100
ccgctcacgc caagcgggaa gatcgatcgg cgcgcccttc cggcgcccga gagccgccgt 26160
cccgatctgg acgtggcgtt tgccaagccc gccacggagc tcgaacgcaa gatcgcgcag 26220
gtctggcaga cgacgctcca gatctcgtcg gtgggattgc acgacaactt cttcgatctg 26280
ggcgggcact cgctcctgct cgcgcaggcg ttcgagcgta ttcggccgct cgtgccgggc 26340
aagcagtggt cgatgatcga gatgttccag tacccgacgc tctattcgct cgccgggttt 26400
ctcagtgagc ggccggacgc cggcgccggg cagcttgccg gcgcgcagga tcgcggacgc 26460
cgccagcgcg agatgctgac gcgtcagggg cctgccctga gcagagtcga aggggtccgg 26520
taatgcagaa cggcatcgcc ctcgtcggga tggccggccg ctttccgggc gcgcggtcga 26580
tcgaggagta ctggcgcaac ctcgtcgccg gcacggagtc gatcacggtc ttcaccaacg 26640
aggagctggc gacggccggg atcagccaga gcgatctgca gaatcccgac tacgtccgcg 26700
cgcgcgggct cgtcccggat ccggacctct tcgacgccgc gttcttcggc atcagccagc 26760
gcgaagccga gctgatggat ccgcagcacc ggctcttcct cgaggcgtgc tggctcgcgc 26820
tcgaacacgc ggcgatcgcg ccgcggtcgt tccccggtct catcggcgtg tggggcggca 26880
tgagcacggg gatgacgaac agcacgtacc tgctgtcgaa tctccacagc cacgcgggat 26940
tgctccagcc ggaggacgtg ctgccggcgc tgctcggcaa cgagaacgac tacctcacga 27000
gccgcgtctc gtacaagctg cacctgcgcg gaccgagcat gaacgtgcag accgcgtgct 27060
cgacgtcgct cgtcgcggtg gtgcaggcgt gtcagtcgct gctcacgtgg cagtgtgatg 27120
cggcgctcgc cggcggcgtg tccgtgtcgt tcccgcaaca ggaaggctac gtctacgtcg 27180
agggcggcat cgggtcgcca gacgggcact gccgcgcgtt cgatgcgaac gcccaaggca 27240
```

```
cggtcttcag caacggcgtc ggcatcgtcg tcctgcgaag gctcgaggac gcgatcgccg 27300
acgggcagac gatctacgcg gtgatccgcg gatcggcgct gaacaacgac ggctcggcga 27360
aggtcagttt cgcagccccg agcgtggacg gccaggcgga ggtcgtggcc atggcgcaga 27420
gcgccgccga cgtggcgccg cagacgatcg actacgtcga ggcgcacggc accggcacgc 27480
gcgccgccga cgtggcgccg cagacgatcg actacgtcga ggcgcacggc accggcacgc 27480
cgctcggaga tccgatcgag gtggctgcgc tcacgaaggc ctttcgcgcc ggcggcgcca 27540
cggagaacgg gttctgcgga ctcggatcgg tcaagagcaa tttcgggcac ctcgactcgg 27600
cggccggcgt cgcggggctg atcaaaacgg cgctcgcgct ccatcaccag cagattccgg 27660
cgacgctgca ctacgtcgag cccaacccga agtgtgattt cgccaacagt ccgttcttcg 27720
tgaacgccgc gctgcgacca tggccgcgcc gtggccatcc gcgacgcgcg gccgtgagct 27780
cgttcgggat cggcgggacc aacgcgcacg tcgtgctcga ggaagcgccg cccgccgcgg 27840
agcgaccggg cgccgagccc gccagcgacc tgctcgtgat ctcggcgaag acgccggcgg 27900
cactcgacgc ggcggctgcg aacctcgcga gtcatctcga ggcgcacccg gacgtcagca 27960
tcgccgacgt ggcgcacacg ctgcgcgccg gccgacagga gtttccgcat cgacgagcgc 28020
tcgcctgcac cagccgcacc gatgcgatcg ccgtgttgcg cggcgcggac gccaaacgtg 28080
tcgccaccgg cgacgccgaa tccggcagcc gtgccgtggc cttcatgttc tccggcggcg 28140
gcacgcagta cgcgaacatg ggccgggagt tgctgccggt cgcgcaggtc ttccgcgaag 28200
aaaccgagcg ctgcctcacg cacctgtcac tctccgccga cgtccgccgt gcgctgtttc 28260
cggatcatgc ggacgtcgag cgggccgcgc attggctcga agctccgtcc gttggtttgc 28320
ccgcgctctt tgtcgtgcag tacgcgctcg cccgtcagtg gatggcgtgg ggtgtgcagc 28380
ccgacgggct catcggtcac agcatgggcg agtacacggc cgcgtgcctt gcgggcgtga 28440
tgtcgctcga gaacgcgctc gcgctcgtgt gtctgagagg ccggctgttc gagacgcttc 28500
cggccgggga gatgctcggc gtcccgcttc ccgaggcaga ggtgcggccc gcgctcggcg 28560
atggactcgc gatcgccgcg gtgaatgcgc caagctcgtg cgtcgtttcc ggcgcgcccg 28620
acgcgctcgc ggcctttgcg cgcgtgatgg ccgaggcggg tgtcgaatgc cgccgcctcc 28680
acatcgcggt cgcggcgcac tcgccgatgg tcgatccgat cctggcggag ttcgagcgtt 28740
tcctccgcgg catcgcgctc gacgcgccaa gggttccagt cgtgtcgaac ctcagcggct 28800
cgtggctgac cgcggacgaa gcgaggtcgc cgcagtactg gacgcggcat ctgagacaaa 28860
ccgtcaggtt cttcgacggg ctcggcacgc tgctcgccga gccgaatcgg gtgctcctcg 28920
aagtcgggcc cggacagacg ctgtccggat tcgcacggca gcatccacag aaaacggcgg 28980
cgcacgtcgt tcaagcgacg ctccggcacg ccaaggaacc ggcgccggac gtcttgttcc 29040
tgcagcaggc ggtgggcagg ctctggacgg caggggtgcc cgtgaactgg ctgcttaca 29100
acagaggcga gggcaggcgt cgtgtgccgc ttccggccta tccgttcgag aagaagcggt 29160
acctggtcga atcgccccag gcgccggtcc cagcgccgct tcaagccctg gcgccaaggg 29220
ccgatgctac tcagcagatg acgacaactg tgccttcgtc tccggcccga tcgcgtgccg 29280
atcgtatccg cgagcggctg accgaaatcc tgcacaaact gagcggcctc ccggcaggcg 29340
agatcgatcc cgcgctgtcg tttctggaga tgggattcga ctcgttgttc ctgacgcagg 29400
cgagtctgcg gttcaaggcg gagttcaagg tccggatcac gttccgccag ctcttcgaag 29460
aggcaccgac gatcgacgcg ctcacgagct acatcgacgc caagctgcct gcggacgcgt 29520
ttcctgcgcc ggcggaagcc gctccggcgc ctttcgtggc cgcaagcgtg ctgtccgcgg 29580
gccctcatgc ggccctggct ggcgtgccgg cggccgcgcc cggcacgctg cagcacgtga 29640
ttcaacagca attgcagttg atggccgagc aactccgcat gctgggcgga aatcccgccc 29700
tgctgcccgc gatgccgccg ccagcgtcgt ccgtcgcggc ggccgacgcg ggactgaagt 29760
cccgcgctac gacggcgccc gtctcgaagc cgcatgcggc cttcaagccg atcgacggca 29820
aagccagtgc ggagctgtcg cccgagcaga tggccgcgct cgccgcctgg atcgatcgat 29880
acacacggcg cacggccgga tcgaagcgac gcctggccga gtaccggccg gtgcttgccg 29940
atccacgctc ggtcgccgga ttcaaccggc tctggaagga gatgatctac ccgatcatca 30000
ccgaccggtc ggaaggatcg aaggtctggg acgtcgacgg caacgagtac gtggatctcg 30060
tgggcgggtt cggctcgctt ctgttcggac accgttcgcc gatcgtggtc gacgccgtgg 30120
```

```
aggaacaact ccaccggggg ttcgagctcg gaccgttgcc gccgctcgcc ggcgaggtcg 30180
cggcgctcgt caaggaattc acggggatgg agcgcgttgg attctccaac accggctccg 30240
aagccgtgct ggcggccacg cgttttgccc gaacggtcac cgggcgcgac aagatcgccg 30300
tcttcgaagg gtcgtatcac ggtctccttg acgaagtgct cagccggccg ctcgtggtga 30360
acggcgaatt gcggtcggcc cccgcggcgc ccggcatcgc cggcagcgcg gtgagcgaag 30420
tcatcctgct cgagtacgcg aacccgaagt cgctcgaggt catccgcgcg cacgagtcgg 30480
agatcgccgc gatcctcgtc gagcccgtgc agagccggcg gctcgacctg cagccgaagg 30540
agttcctgca cgaggttcgc cgcgtggcgg acgagatcgg cgccgcgctg atcttcgacg 30600
aagtcatcac cggcttccgc gtgcatcccg gcggcgccca ggcgcacttc ggcgtccgcg 30660
cggatcttgc cgcctacggc aaggtcgtcg gcggcggcat gcccatcggc atcgtcgcgg 30720
gcaatgcgaa gttcatggac gcgttcgacg gcggccactg ggagtacggc gacgggtcgt 30780
tcccggaggt gggggtcacg tactttgccg gcacgttcat gcggcatccg ctcgcgctcg 30840
ccgctgcgaa ggccgtgctg acgtacatga agtcgcaggg tcccggcctg cagacgcggc 30900
tggcagatcg gatcgagcgg ctcgcgaacg acgcgcgcgc gatcgtggcc aggcacggcg 30960
cgccgtatca gattacgcag ttcagctcga tgatgagcct gaacttcccg cacgaccaga 31020
agacggcggt gctgctgttc ttcctgatgc gcgagcgcgg gattcatatc tgggagggcc 31080
gtcccttctt cttcacgacg gctcacaccg aggccgacta cgccgccatc ctccgcgcgc 31140
tcgacgagtc gctggccgag atgcaggcgg ctggcttctt cggcgcgccc gcccactccg 31200
cattgcaggt gaccgggacg gcggaccagg acgtcctgcc gttcaccgaa ggacagcagg 31260
agatctggct cggcggccag atgggcgagg cggcatcacg cgcctacaac gaagtcgtcg 31320
cgctcgacct cgcgcgggcct ctgaatcgcg cggcgatgca gcgcgccgtc gacgaggtcg 31380
tcgcgcgtca cgaatcgctg cgcatgacgg tggccagcga accgctcggc ctgcgtcacg 31440
ttccgggcac gtcggtgccg gttgaatggc aggacgtgtc gggattgccg gaggacgctc 31500
agcacgaggc cgtgcagaag gtgctcgagg cggcgacgg cgtggcgttt gacgtggaac 31560
gcgggccgtt gctgcgcgtg accaccatga cgctcacccc cgagcatcac gtcttcgtca 31620
tggcggccca ccatctcgtg tgtgacggct ggtcgttcgg cgtcatcctg cgcgatctgg 31680
cgacgttcta ttcggcccag gttcgaggca cgcgtgcgaa cctcgaggcg ccgatgcagg 31740
tgagccgctt cgcgcgtgag gatcgcgagg ccaagcagag tcctgaagcg gccgagaccg 31800
aggcgttctg gatccggcaa ttcgacaccg tccccgagcc gctcgagtta cccgccgatc 31860
ggccgcggcc gccagtgaag tcgtaccagg gcgcccgtgt gtccgtgccg tttgacgcgg 31920
cgctcgctcg cggcgccaag aagctcgccg ccgaacaccg cacgaccctg tttacgacgc 31980
tgctcggcgt cttcgagacg ctcgtctacc gactgaccgg ccagcaggat ttcgtcgtcg 32040
gcatcccggt cgcggtgcag cctctgctcg gcgaggatct ggtcgcgcac tgcgtcaact 32100
tcgtgccgct tcgcgcgcgc gtgtcaccca ccgcgacgtt ctcggagtac ctgaacacgc 32160
tgaggacgca gtcactcgac ctcaatgagc accagaactt cacgtacggc agtttgctcg 32220
ccaagctgcg gttgccgaag cacccgggcc gcgatccgct cgtctcgacg agcttcacgc 32280
tcgagcccgc catcgtcgat cccgggttcg acggtctcac ggcgcgttcg ctgacggttc 32340
cgcgcgtgac atcgaagcgc gacctgcacg tgaacgtcat ggagatggac ggcggtctgc 32400
tggtcgaggc cgactacagc accgatctgt tcgacgaacc gacggtgcgc cggtggctgg 32460
atagctatcg gatcctgctc gaggccgtcg tgtcgtcccc cggccggtcc ctcctgagcg 32520
tgccggtgat ctcggagccg gaccggagtc agcttgtcac cggctggaac gacacggccg 32580
ccgactaccc gcgcgatcgc gtcatccatc aactcgttga agagcgcgcc gtgcagaccc 32640
cggagggact cgcgctcgtc tgtggtgccg agcggctgac ctggaccgag ctcaaccgac 32700
gcgcaaaccg gttggcgcat cgcgctcgtga agaagggcgt cgccccgggg agcagggtgg 32760
cgctgtgcct ggaacgatcg gccgggttca tcgtctcggt gcttgccgtg cacaaggccg 32820
gcgccgcata tgtaccactc gacccgtct cgcccggcga tcgcaagtcc ttcgtcgtcg 32880
aggattccgg tgccgtgctg gtcgtgaccg attcgcggtc cgaaggccga tggctgacgc 32940
ccaggacgcc cgtgctccac ctcgacgccg acagtccgcg aatcgcgcaa gagtcgcacg 33000
```

```
atgatctgaa ggtcgccctg tcggccgaag acgccgcgta cgtgatctac acgtccggat 33060
cgacgggaca accgaagggc gtcgtcatcg ggcaccgcca gctcgtgaat tacacgtggg 33120
cggtgatcga acggatcggg ttacccgccg gctcgagtta tgcgctcgtg tcgaacgtcg 33180
cggccgatag cggtgtgacc gggctctcgg cgctcgcgat gggctgggtg ctgcacgtga 33240
tcacctcaga ggtcgcgacc gacgggcggg cgctcgggtc gtatttgacc gttcacgaca 33300
tcgacttcct gaagatcacg ccgtcgctgc tcgcgtcgct gctcggcgat cacccgtcgg 33360
cctcgctgct gccgcgccgc tgtctcatgt tgggcggcga gccgtcacgt ccggcgtggg 33420
tggaagagct gcagcgtctc gcaccaggct gccggatcat gaaccactac ggcccgacgg 33480
agaccacggt cggcgtgttg acgtactgga ccggtgaccc cagggatctg ttcgtcgaga 33540
gcgtgccggc cggcaggccg ctcgggaacg tccgcgccta cgtgctcgat ccggccggag 33600
cgcccgcgcc gattggcgtc gtcggcgagc tctgtatcgg cggcgcgtgc gtggcacgcg 33660
ggtatctcaa tcgtcccgac ctcaccgccg agcgcttcgt ccccgatccg ttcgccaccg 33720
agccgggagc gcgcatgtac cggaccggcg accgcgcccg gttccgggcc gacgggaaca 33780
tcgagttcct cggtcgcgcg gactatcagg taaaagtacg cggattccgc gtcgagctcg 33840
gcgaaatcga agcggcgctt cgcgctcatg agggcgtcga gcaggccgta gtcgtgctgc 33900
gaaaagatca gcccggcgac gagcgcctga tcggctatgt cacgaccggc ggcggcgcct 33960
cgatctcgat ggcggagctt cgcacgtcgc tcaagcagag gttgccggac tacatggtgc 34020
cggcggtgat cgtcgcgctc gacaagctgc cggtcacgtc gattggcaag atcgatcgac 34080
gcgccctgcc ggcgccgccc gagcgcgcca gtttcgagtc ggagttcgtc acgcccgcca 34140
ccgagacgga cgcccggctt gcggagatct ggagcgcggt gctcggtctc gagaagattg 34200
gcgcgcacga caacttcttc gacctgggcg ggcactcgct gctcgcggcc cgcaccgcca 34260
tgaaggtccg cgacacgttt cgcgtggacg tgtcgctcat cacgttcttc gaaatgccca 34320
ccatcgccat gatggcggca gcaatcgacg ccggtaccgc ggcccctgcc tcccggacca 34380
tcacgccgag ggcccgcgtc agggcccgca gggatgagct ggtccgccca tgaaccccgg 34440
ccacccgcag actgagtccc ccgacctgcg gcgcgacggg atcgccgccg aagcgggcga 34500
cgtgtttctg gcgccggcgt cgtacgccca ggggcgcttg tggttcctgg cgcagctcga 34560
gcctgaaagc tccgcctaca acattccgat cgcgctcaga ctgaagggcc gcctcgaccc 34620
cgccgcgctc gaacgcgcca tcaacaccat cgtccagcgt cacgagacgc tgcggacgac 34680
gttctcgctg cagggcgatg agctcaagca ggtcatctct ccgacgctgc gtattccgat 34740
ctccctcgtg gacatgcgtc tgctcagtcc tgagaaacgc gaggccggat ggcgccgcgt 34800
tgcggccgac gacggcgcgc gcgtcttcga tctggctgag ggcccgctgc tcaaggtcgt 34860
actcgtcatc ctcgagccct ccgatttcgt cctcctgatc accgtgcatc acatcgtctt 34920
cgacgggtgg tccggcggga tcttcatccg cgagctggcc gaggcgtacg cggctgagct 34980
cgacgggcgc gacccgaagc tgccggagct gccgatccag tacgcggact ttgccgcgtt 35040
ccagaaggaa cagaccgaag ggcccggcct gccggtgctg ttgaagcact ggacgcagga 35100
gctggcgggg gcgtcgatga cgctcgacct gccgacggat cgtccgcgtc cgcccctgca 35160
gacctcccgg ggcgccctcg cctcgcttcg gctcgacgga gcgatctggc cgggcgtcgc 35220
cgcattcagc cggggcgaga acgcgacggt gttcatgacg cttctctccg ccttcggcgc 35280
cgtgctgcat cgctggtccg gacaggaaga catcctcatc ggcagccgt cggccggccg 35340
cgatcgtccg gagcttgagg gcctgatcgg gttcttcatc aacacgttcg tcctgcggct 35400
caacgcgtcg ggagatccca cgtttcgcga gttcgtcggc cgggcgcgcc gcgcgtgccg 35460
tgcggcctac gcgcatcagg acctgccgtt cgatcggctc gtcgagacgc tcaacccgga 35520
gcggacccgg gatcggcatc cgatctatca agtcatgttc gcgcaggatc cgccggcgtc 35580
gaagatgacc ttcgccggga tcgagctcga gcggcttgtc gtccacaacc gcagcaccaa 35640
gtgcgacctc acgctcgagg ttgccgaaga cgccgacggc gtcaccctgt acgtcgagta 35700
cagcgtcgac ctgttcgatg cggcgacggt ggcgtcgctg ctgcagcaat atcgccgtgt 35760
gcttcagacc gcgatcgcgg acccggacca gcggctctcc gagctgcgtc tgcttggcga 35820
cgacgagcgc atgcagctga tcgcactggg gaccggtccg gcggccgtct acccgcgcga 35880
```

```
cgtcagtctg gccacggtct tcgaagagcg ggtccgcgcg actccggcgg cagtggccgc 35940
gacgctggag ggacagcacc tcacgtacgc ggatctcaac cggcaggcga accgcctggc 36000
ccggcggttg aaggcgttcg gtgtgggccc gaacgtgctg gtcggcgtgg cgctcgaccg 36060
gtcgttcgat ctgctggtgg cgctcctcgc cgtggcgaag gccggcggcg cgtatctccc 36120
gctcgatttc gctcacccgc aggaacggct tgcgttcatg ctcgccgaca cgcgcgcgcc 36180
catcgtgctc acgctgcggc gatttgccgg ggcgctcgag tcctttcccc tcttctgcct 36240
ggacgacgag ctgccggcca ccgccgagga gccggacgaa gatctgccgc cgcagtcgac 36300
cggagaccac ctggcctatg tgatgtacac ctcgggctcg actggtcagc caagggcgt 36360
cgcgatcacg catcgcggcg tggttcgtct cgtacgcggc acagactacg tccgctggaa 36420
cgacgtgcgg gcgatcctcc acatcgcgcc gacgtcgttc gacgcgttga cgttcgaggt 36480
gtggggcgcg ttgctgaacg gcgcgcggct ggtgctggtt cccgagcggc tcgtcagcct 36540
ggagacgctc gagcgcacgc tccggtcgga gcaggtggac tgcctgctgc tcacgaccgc 36600
gctcttcaac gccgtcatcg acgcgaagcc cggcattctc gtcggcgtga agcaactcct 36660
gatcggcggc gaagcgctgt cagtcgcgca cgtgcggcgc gccctggcgc acctgccgga 36720
cacgcggctc gtgaacgcgt acggtccgac cgagtgcacg accatcgcgt gcgcctacca 36780
gattccgcgc acgctcgatc cgcaggcgcg gtccatttcg atcggccggc cgattgccaa 36840
tacgcaggcg ctcattctcg atcgacacgg cgacctcgtg ccgttcggca tcgccggcga 36900
gctctgcctc ggcggagacg ggctggcgcg ggaatacctg aatcagccgg cgctcactgc 36960
ggaacggttc gtgcccaacg cgttcagccc cgagccggcc gccggctgt accgcacggg 37020
agacctcgcg cggatgcgcc gcgacggcaa catcgagttc cttggccgga tggatcgcca 37080
gatcaaactg cgcggcttcc ggatcgagcc gggcgagatc gagcatgccc ttcgtgcccc 37140
gggagacgtt ctggacgcgg tggtggagat ccgggtcgat gcccacatgg cccatctcgt 37200
cgcctacgtc gtgcgcgccg atgggagtca actcaccgga accgacctgc gcgaacgtct 37260
gaaggcgcgc ctgccggaat actgtgtgcc cgcgaagtac gtgttcctgg ccgaggtccc 37320
gcgaacggcg gccggcaagg tcgatcgctc ggcgctgcct gcatcggcga tcgaggcgcc 37380
cgcgcccgag ccgcgcgatc cgtcgttcct gaacgaggtc gaagcggagc tcgcgcgaat 37440
ctggagcgaa gtgcttcggc tgccctccgt gctcgcgacc gacaatTTTT tcgatcgtgg 37500
```

<210> 2
<211> 37507
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: DNA sequence
      of clone FS3-135.

<400> 2
```
gatcccacga tcggcgccag cggattgcgc agctcgtgcg cgagcatcgc gaggaactcg 60
gtcacgtggc ggccctcggc ctccagctcc tccatgcgct tgcggtgcga gaggtcgcgc 120
gtcaccttgg cgaagcccat gtgccttcca ctcgcgtcgc gaagcgccgt cacgacgacg 180
tcggcccaga agcgcgtgcc gtccttgcgc acgcgccagc ccgtgtcctc ggcgcggcgg 240
tgcagcagcg catcgtgcag ctcggcctcc gggcggccgg ccgcggcatc ctcgggcgga 300
tagaagagcg aaaagtgccg tccgatgatc tcttcggcgc tgtagccggt gatgcgctcg 360
gcgcccttgt tccagctggt cacgtggccc gcggggtcca gcatgtagat cgcgtagtcc 420
tgcacgccct cgatcaggag gcgcgagcgc tcctcgcttt gccgcagctg ctcctcgtgc 480
tcgcggcgct cggtcaggtc gcgcgtgacc ttgccgaagc cgaccagcgt gccgtcctcg 540
```

```
tcgcgcagcg ccgtgatgat gacgttggcc cagaagcgcc ggccgtcctt gcgcacgcgc 600
cagccctcgt cctcgaagcg gccgtcctcg cgtgcgcgca gcagctcctt ggcgggccag 660
tcgcgcgcga tgtcctcctc ggtgtagaac gcggagaaat ggcggccgat gatctcttcg 720
gcgcgatagc ccttcaactg ttcggcgccc gcgttccacg aacgcacgta gccctcggga 780
tcgagcacga agatcgcgta gtccgtgacg gacccgacca gcagccggaa gatggcgtcc 840
tcgccggcca gcccgctcgg ggcgccgaag gcgaggttgc gcaccgaggg ccgcggaccg 900
gatgggcttg gcatggctat tagggtacaa cggtgcacgc aagttgcaat cggacaatca 960
gagataattg cttgatgaag atgccatttt tggaattctt acgcagggag ctggggacgc 1020
tcgaggccca gggcctgtac aagtccgagc gcgtgctcga ctctccgcag ggcgccagcg 1080
tgcaggtggc cgcccacgac gtgctcaact tctgcgccaa taactacctc ggcctggcca 1140
accacccggc cctggtcgcg gccgcgaagg agacgctcga gcgcgacggc tacggcatgg 1200
cgtcggttcg cttcatctgc gggacgcacg ccgtgcaccg cgagctggag acgcgcctcg 1260
cgcgcttcct cggcaccgag gacgcgatcc tgtacggctc ctgcttcgac gccaacggcg 1320
ggctcttcga gacgctgctg ggcgaggagg acgcggtgat ctcggacgcg ctgaaccacg 1380
cctcgatcat cgacggcgtg cgcctctcga aggcgaagcg cttccgctac gacaacaacg 1440
acatggcgag cctcgaggcc cagctgaagg cggcggacgc ggccggcgcg cgctacaaga 1500
tgatcgccac cgacggcgtg ttctcgatgg acggcatcgt cgcgaacctg ggcgccatct 1560
gcgagctggc cgggcgctac cacgcgatgg tgatggtgga cgactcgcac gccgtgggct 1620
tcatgggcgc gcacgggcgg ggaacgcccg agcattgcgg tgtcgaaggg aaggtggaca 1680
tcctcaccgg cacgctcggc aaggccctgg gcggcgcttc cggcggctac acggcgggca 1740
agcgcgaggt ggtggcctgg cttcgcaacc gctcgcgccc ctatctcttc ccaacacgc 1800
tgatgcccgc catcgcgggc gcgtcgctca aggtgctcga tctcctcgag ggcggcggcg 1860
agctgcgcgc gaagctcgcg cgcaacgccc gccacttccg cggcgagatg acgcgcctcg 1920
gcttcacgct ggccggcgcc gaccatccga tcatcccggt gatgctgggc gaggccccgc 1980
tcgcgaagga gatggcggac cggctgctga aggaaggcat ctacgtggtc ggcttctcgt 2040
ttcccgtggt gccccgaggg caggcgcgca tccgcacgca gatgtcggcc gcccatgaac 2100
cgaagcacgt cgatcgcgcc atcgccgcct ttgccaaagt cgggcgcgat ttgggagtca 2160
ttgcttgaga accctttcga agacgaagcg cgagcccggc atctggatgg tcgagtcgcc 2220
caagcccgtg gtcggccaca acgacgtgct gatccgcgtg aagaagaccg ccatctgcgg 2280
caccgacatg cacatcttca actgggacga ctggtcgcag aagaccatcc cggtgccgat 2340
gacggtcggc cacgagtacg taggcatggt ggaggccatg ggccaggagg tgcgcggcct 2400
gcaggtcggc cagcgcgtct ccggcgaggg ccacatcgtc tgcggccatt gccgcaactg 2460
ccgcgccggg cgccgccacc tgtgccgcaa cacgcagggc gtgggcgtga accgccccgg 2520
cgcgttcgcg gactacctgg tgattcccgc ggagaacgcg ttcccgattc ccgacgacat 2580
ccccgacgag atcgcctcga tcctcgatcc gttcggcaac gcggcgcaca ccgcgctctc 2640
gttcgacctg gtgggcgagg acgtgctgat caccggcgcc gggccgatcg gcatcatggc 2700
cgcggccatc gcgccacg tcggcgcgcg ccacgtcgtg atcacggaca tgaacgacta 2760
ccgcctggcg ctcgccacga agatgggtgc cagccgcgcg gtgaacgtat cgaaggaaaa 2820
cctgaaggac gtgatgcgcg agctcggcat ggtcgagggc ttcgacgtcg gcatggagat 2880
gtcgggcgtg ccctcggcct ccgccagat gctcgacacc atgaaccacg gcggcaagat 2940
cgccatgctc ggcatcccgc ccagcgaggc cgcgatcgac tggacccagg tgatcttcaa 3000
gggcctggtc atcaagggcg tctacggcg cgagatgttc gagacctggt acaagatgat 3060
cgccatgctg cagagtggcc tcgacctctc gcccatggtc acgcaccgct tcgacgtgcg 3120
cgagtacctg aagggcttcg agacgatggg ctccgggaaa tccgggaaag tcgtcctgtc 3180
gtgggattag cgagggccac ggatcgcggc ccggttcccg cggcggcgat cgtgcaggcg 3240
ggttggcagc acgccccggg atttccgccg tggctcacgt gcgacttcgc cgtcgacgca 3300
ccgatcgcga tcgacgtcga cgtgagcgtg tcgctgctcg atgcccaggc cacctcgctg 3360
tggacgcagt cgatggccgg cagcgggcgc gacaccccgt ggctaccgac cggcacgtat 3420
```

```
cgcgcctccc tgcgcctgga gccgttccgc ttcccggccg cggccgcctc ggtcgagttc 3480
gcgctctcga tgcgggtgca aggtgagcgc cgcgtcgtgg ccaccgcagc gggcgtggtg 3540
cccgcgggcg caccccacgg cgtggcgcgc cggcctggc acctcgaggc cctggatggc 3600
acgccggcgc tcgagacgct ggcctggtcc gatcccacgc acagctggtt ctcccagcat 3660
ttcgaccatg cgtcgcgcac gtgcgtcgag tacctgggcg ccggatggcc cgggtggcgc 3720
gggcgcgtgc tggatgcggg atgcggcgac ggcatcacgg cgctcggcat cgcgctgcgc 3780
tatgcgcccg agcaggtggt gggcatcgat ccggggcgct gctaccgggt gcttcccgac 3840
atcctcgagc gccacgccct cggccacctc gcgctgccgg acaacctgca gttcctcccc 3900
gcggatgcca acgcgctccc gttcgccgac agcagcttcg acgtcatcgt ctcgtggtcc 3960
gcggtcgagc atttcgtggg cggctacctg ccctcgctcg ccgaggcgcg cgggtgctc 4020
aagcccggcg gcctgctggt catccatccc gagctctact acaccgcgca ccacggccat 4080
cacctcggcg agtacagccg cgagcccttc ttccacctgg tgaagtcgcc cgacgaggtg 4140
cgcgacatcg tgttcggcgc ggacgtgagc ctccacgacc gcggcggcat cgcgccgacg 4200
cgcgcggagc actggcgctg gtacaacgag ctcaatcgca tcacggccgc gcagttcgag 4260
gacgagttgc gcgcgctcga cttcgagccg tggcgcctgg cactgcgcgc cgagccgctg 4320
gtggagtaca agccggagct gctgcgctac cgcttcacgg agctgggcgt ctccgagctc 4380
tacgtggcct gcatcaaccg caagcccgtg tcagatggtt cctacaggac gcatgggcag 4440
gcggcattcc ggtaaggtcc gccgggacgg gaattgaagg cgcaagcgcg ttgttatccc 4500
cgtcactccc ggactgccgg gatcaccctc acaggagaac ccccgatgaa gacccagacc 4560
caggccgtgc tcgccgcggc cctgctgtcc cttgccggca ccacgctggc ccgcgccgaa 4620
gccgtcacgc tgacgggcgc gagcgaagtc ccgcccgtca cctcgagcgc gaagggcagc 4680
ggcacggtgg tcgtgaaggc cgattgcacg gtcaccgcga agatcaccgt caccggcatg 4740
accgccaccg cggcccacat ccacgaaggc aaggccggcg ccaacgcggg cgtggcagtc 4800
ccgttcgtga agaccgcgga cacacggttc gaggcagccc cgggcgccaa gatgaccgag 4860
gcgcagtgcg ccgcgtacaa ggcgggcggc acctacgtga acgtgcacag cgaggcccac 4920
aagggcggcg aggtccgcgc ccagctggcc ggcaagtagg gcagccgctg caaaaagaaa 4980
aagcccgcgc aagcgggctt tttcgttgaa ggctttggtc ggggcgaaag gattcgaacc 5040
ttcgaccccc tgcaccccat gcaggtgcgc taccaggctg cgctacgccc cgaccgaact 5100
tgaaattata cctgcatcgg ggtcgcagcc gccatgaatc gaaacgcccc gggcttcggg 5160
tagcatccgc ccatggcgac gaccctcgtg acgggtggtg caggctacat cggcacccac 5220
atcctctgcg cactggccca ggccgggcga cgcagcatct gcatcgacaa ctactccaac 5280
agctcgccgc gttccatcga acgcgtgctg cagatcgcac ccggttgcgt cgaggccttc 5340
gacgtggaca tccgcgacgc cgacggcatc cgcaaggtga tcgcagggcg cgacgtggac 5400
agcgtgatcc acctcgcggg actgaaggcg gtcggcgaat cggtcgagca gcccgagcgc 5460
taccacgaca acaacgtccg cggcaccgag agcctgctcg cggcgctcgc cgactcgccg 5520
gtgcgcaagt tcgtcttcag ctcctcggcc accgtgtacg gcctcgccga gaagatgccg 5580
atcgacgagg acgcgcccac ctcgccgcag agcccgtacg gccagaacaa gctcgacatc 5640
gagcacatgc tggtcgccct cgccaagcgc gaccctcct ggcgcgtagt gaacctgcgc 5700
tacttcaacc cggtcggcgc acacgagagc gcgctgatcg gcgaggaccc ggccggggtt 5760
cccaacaacc tcatgcccta cgtgtgccag gtcgcctccg ggcgactgaa ggagctgagc 5820
gtctacggca gcgactaccc caccccggac ggcaccggca tgcgcgattt catccacgtc 5880
tgcgacctgg ccgaaggcca cgtggccgcg ctggaagcac tcgagcgcgc ggccccgggg 5940
accgtcctca cggtgaatct cgggacgggg ctcggctaca ccgtgctgca attgctcgaa 6000
accttcgagc gcgtgaacaa actaagcatt gcccggcgca tcgtggggcg cgcggccggc 6060
gacgtcgcga tctgctacgc gaacgcgggc cgcgcgaagg ccgtgctggg ttggaccgcg 6120
aagcgcggaa tcgaggagat gtgccgcgat gcctggcgtt ggcaggaaag gaatgcaaag 6180
ggctacggtt gacggattgc aaatcatggc cttggcgacc cttgctgcaa tgcggcatga 6240
aggggcggt tcttgtaaga tccaaggttc gggatgcgtc caggacgtca aaaacagcac 6300
```

```
ccggcgttgg aaggaagaag gcaaccctag tgttgtcatg gctccagacg ggattgcgcg 6360
ccgagtaaga atcgtcagtc gagaccaact cgcctccggg tgcaacgaat cgtgagccga 6420
agcaacacct cgcatcgcgt cgccgtcgtc ggcattgcgt tccgccttcc cggcgctacg 6480
cgcgaatccc tctggccctc gctcctcgag gggcgcaacc tcgtcaccac cgtcgatccc 6540
gcgcgctggg cgaaggagcc gtactaccac ccgcgcaagt ccgagcccgg caccagctac 6600
tcgttcgccg ccggcaccat cggcgacatc agcggcttcg acgcgtcgtt cttcggcatc 6660
tccccgcgcg aggccgcgca gatggatccg cagcagcgcc tgctgctcga gctctcctgg 6720
gaagccctcg aggccgcggc ggtgaagccc tcgaccctgc gcggcagcga ttgcggcgtg 6780
ttcgtgggcc tgtccaccgt cgactacagc tacggcctcg cctacgacct ggcgtccatg 6840
gacgcgtcgt cggccaccgg caacaccgcg agcatcgcgg ccaaccgcct ctcgtatttc 6900
tacgacctcc acggccccag catggtcgtg gatacggcgt gctcgtcctc gctggtggcg 6960
ttccaccagg cctgccagtc gatcgccatc ggcgagacgc gccaggcgct ggtgggcggc 7020
atcagcctgc acctgcatcc gtacgggttc atcgcgttct ccaaggcgtc gatgctctcg 7080
aagcgcggcg cctgccgcgt gttcgacgcc agcggcgatg gctacgtgcg ctcggaaggc 7140
gcgggcgtgc tggtgctgaa ggacctcgac cgcgccgtgg ccgacggcaa cccgatcctc 7200
gccgtcgtcg ccgcgagcgg cgtgaatacc gcgggccgct cctccagcct caccgtgccc 7260
agcgtcgatg cccaggctgc attgctgtcg gatgtgtacg agcgcgccgg catcgacccg 7320
tccgagatcg actacctcga ggcccacggc accggcaccg tggtgggcga tccgatcgag 7380
acccgcgccc tcggcatcgc gctcgggcgc ctgcgcccgc gccaccagcc gctgccgatc 7440
ggttcggtga agagcaacgt cggccacctc gaggcggcct cgggcatggc gggcctggtg 7500
aaggcgctgc attgcctcaa gcaccgcgcg gtgccgccca ccatccacct cgattcgccc 7560
aaccccaaca tccacttcga cgagtggaac ctgaagccgg tcctcgagcc gctcacgctc 7620
catccgtaca agaagctcgt cgtcggcgtg aactcgttcg gcttcggcgg cgccaacgcg 7680
cacgtgatcc tcgagagcac gcccacttcg cccaaggcgg tcgaggcccc ggccggcgag 7740
gtggcgctca tgctgagcgg ccacgacgac gcggcgctgc gcgacgtggc ggcctcgtac 7800
gcgaacttcc tgcaggcgag cgagcaaccc gcgctgcacg acatcgcgta caccaccgtg 7860
cactgccgcg actggcacca gctccgcgcc atggtggtcg gcaccgaccg cgccagcatc 7920
gccggcgccc tcgcgcaatt cgccaagggc cacaacacgc ccggcgtcgt gagcggccgt 7980
gcgctgccca acgcgtccgg ccccgcattc gtgtattcgg gcaacggcgc gcaatggccg 8040
gcaatggggc gcgccctgct cgccggcgag ccggtgttcc gtgccgccgt gcaatcggtc 8100
gacgcgctgt tccgcccgct cggcggctac tccatcctcg agatgctgga gcaggaggtc 8160
cccgccgacc agctggaccg caccgagatc gcccagccgc tgctcttcgc cgtgcaggtg 8220
ggcttgaccg agctgctgcg ccactggggc atccagccca cgcgcgtcac cggccacagc 8280
gtgggcgagg tcgccgccgc ttgggcgagc ggcgcgctga cgctcgacga tgcggcgcgc 8340
gtgatctacc accgcagccg ctaccagggc gagaccaagg cgcggggcgc gatgaccgcc 8400
gtcggcctcg ccgaggacgc cgcgcgccgc cagatcgagg cgctcgggtt gaagggcgag 8460
gtgaacgtcg cgtgcatcaa cagcacgcgc aacgtgacgc tcgccggcac gcgcgccggc 8520
atcgaggcgc tcgaggccga gctcacgcag cgcaaggtgt tccaccgccg cctcgacctc 8580
gactacgcgt tccacagccc ggcgatggac ccggtgcgcg acgggctcgt gagcgcgctg 8640
cgcggcctca cgccgcgcgc caccgggtc gcgttccatt cggcggtcac cggctcgccc 8700
gccgcgggca accagctcga cgccacctac tggtggagca acatccgcga gccggtgcgc 8760
ttccaggccg cgatccgcgg catcgccgag tcgggcgtca atgtcttcat cgagatcggc 8820
ccgcacccga tcctcaagaa ctacatcaac gacggcctgc gcgccgcctc gatcgagggc 8880
cgcgcgctgg tcacgctgca gcgctcggcg agcgatcgcg cggcgatccg cgccgccgcc 8940
caggaagtcg tgatcaccgg ctgcccggtg gagaccgcga agatgctccc cgcgcagggc 9000
cacttcgtcg agctgccgcc ctacccgtgg cagcgcgagc gccactggcg cgcgcccacc 9060
tcgcaggcct acgacctcat ccagcacggc aagcagcacc cgctgctggg ctatcgcctg 9120
cacgagaacg acttccagtg ggaaaaccac atcgacaccg cgctctatcc ggcgtacgcg 9180
```

```
gaccatgtcg tcggcggcgc cgtggtgttc cccgcggccg gcttcgtcga gatggcgctc 9240
gccgcctcgg cgatcgcgct gggcggcgac gcgcacgaga tcgagacgct cgagatccgc 9300
agcgcgctgc tgctcgagga ctccacgtcg aagaccgtgc gcttcgcgct ggaacccgat 9360
ggccgcttca caatccgcag ccgcgctcgc ctgagcgagg atccgtggca actccacgtg 9420
gtcggcaagc tggtcggcac gcccaccgag cttccgcgcg cgccgcgcca gtcggcgccg 9480
tcgcgcaagc ccgatgtgat cgcggcccag cactacgaag gcgccgcgaa ggccggcctc 9540
gcctacggcc cggccttcca gtcggtctcc aaggtgtggc tcgacgccgc ggacccgggc 9600
agcgcctacg cgcgcctgat cctccccaag ccgattcgcg gcgagctggg cgtgatgcac 9660
ctgcaccccg cgtcgctgga cggctgcttc cagctgctgg tcgacctgct gcgcgccgag 9720
gcctcccggc acgcgcagat cgccttcgtg ccgatcctgg tcgggcgcac gcgcctctac 9780
ggccctgccg ccctcgtcac cgcctgccgc gtccgcctga ccgcgcgcag cccgcgctcg 9840
ctggtcgcgg acttcgagct ctacggcccc aacggcgacg tggtggccgc gctcaccggc 9900
gtgcgcttcc gcggcgtcca gctgaagcag ccgcagggcg cgcgcctgcg ctacctgtcg 9960
taccgcgcca tcccgcgcct gaatggcggc gacacgccgg cgccaagcc gctgcccctc 10020
gcgttgctgg ccgcggcgtg cggcgagcgc tggcattcgc tctcggccgg ctcgcgcaag 10080
cgctactacg atgaggtcga gccgctctcc gacgtgctgt gcagctcctt cgccgagcac 10140
gcgctgcgaa agcggatcgg cgaggccccg ctcatcgacc cggccgcgct cgccgcgtcg 10200
acgccggccg agcattggcc gctcctcaag cgcgcgatcc agatgctggt cgaggaccag 10260
ttgctcgagc ccgccgaagg cggctggcgc tggagcccgg tcaccgaatt gcccgacgcg 10320
caggagagct ggatcacgct gctgcgcgac tacccggacg aggccggcca gctgatcgcg 10380
ctcgggcgcg ccggcgctca cctcgcggaa gtgttcgccg gcaaggaagc gtccttgccg 10440
agcctgttgc cggccacgca ggcgaacgtg ttcgcgcgcc tgtgctcggg cgcacccgcg 10500
ttcgccaacg ccgggcgcgc catcgccgac acgctggcgc tcgccctcga gcgcctgccc 10560
gcgcaccggc cgctgcgcgt gctcgaggtg acgccgtgcc gctccgagat cacgctgcgc 10620
gtcctcgagg ccgtggacct cgaccgctgc gagtaccacg tcggctgcac cacggacgaa 10680
gcactgggcg agtacgaagg cgtcctcgat gccattccgc acgtcgagac ttgcgtcgtg 10740
gacctgaaga tcccgggcat gggcctgaag gcgccgaacc agggcccgtt cgatgtcgtg 10800
atcgtttccg atggcctgct cggtgcggcc gatccggatg cggccgcggc gcacctcgcc 10860
ggcgcgctcg cggaagacgg gctgctggtc atggtcgcgc agcatccgtc gcgctggacc 10920
gacctcctct tcggcgcgga cccggcgtgg tggacgctcg ggccctcggg ctcgcagcgc 10980
tcgcgcctgc gttcgccgga cgagtggcgc atcgcgctgg cccacctcgg cttcgggccc 11040
accgtcgcgg tgcccgagct gcccggcctc gagcgcggct cgtacgtcct catggcgcgg 11100
cgcgacggcc cggtgccgca ggctgccgag cgcagcgttc cgccaagtc gggctggatc 11160
ttcgtgcagg aagcgggcgg ttattcggcg gcgttcggcg cctgcgtgca ggcggatctg 11220
cgcgcacgcg gccacgcggt gttcgagatg gcgcccgatt ccgtcgaggg ggccgagatg 11280
gcgcagcttg cgcgcgcgcg ccagacgctg ggcgacatca gcggcatcgc cttcctggcc 11340
ggcctgcccg acgacgcggc cgccacgggc gacgcggccg agcgcctcgc gcgccagacc 11400
gcgcgctgcg ccgcgctcgc gcgcctgctg cgcgaatgca aggcgagtgc catgacaccc 11460
gcggtgtgga tcgtgacctc gggcgcggcc tccagcctgg cgcgcaacgt ccccgccctt 11520
gcctcgcgcc gccgccctcc cgatccggac gaggcgatcg tatgggggtt cgcgcgcacc 11580
gcgatgaacg aattccccga gctgcgcctg cgcctcgtgg actgcccgaa cccggagaac 11640
ctcgagcgca acgccgcctc gctggtgcag gagctgctct cgcccaccgc cgacgacgag 11700
gtggtcctca ccggcgccgg ccgctacgcg ctgcgcctgg cgcggtgcc gccgcccgca 11760
tccgtgcgca ccgctaccca gggcatcacg cgcctcgact tcgccgcgcc cggcccgctg 11820
aagaacctcg cgtggcgcgg cgacgtgcgc cgcaagccgc gcgcgaacga ggtggaaatc 11880
gaggtgcgcg ccgcgggcct caacttccgc gacgtgatgt acgcgatggg cctgctctcc 11940
gacgaggcgg tcgagggcgg cttcgccggc gcctcgctcg gcatggagct ctcgggcgtc 12000
gtcaccgccg tcggcaagga cgtgacctcg gtcgcgcgcg gcgacgaggt gctcgcgttc 12060
```

```
gcgccctcgg cgttcgccac gcacgtggtc accaccgccg attcggtcgc gaagaagccc 12120
gccggctgga cgttcgagtc cgcggccacc atcccgaccg cgttcttcac cgtgtactac 12180
tcgttgaagc acctggccca gctgcgcgag ggcgagcgcg tcctgatcca cggcgccgcg 12240
ggcggcgtgg gcatcgccgc gatccaggtc gccaagtggc tgggcgcgga gatcttcgcc 12300
accgccggct ccgacgagaa gcgcgacttc gtgcgcctgc tgggcgcgga ccacgtgctc 12360
gactcgcgca cgctcacgtt cgccgacgac gtgctgcgca tcaccaacgg cgagggcgtc 12420
gacgtggtgc tgaactcgct ctccggcgag gccattgcgc gcaacctgcg cgcgctgcgc 12480
ccgttcggcc gcttcatcga gctcggcaag cgcgactact acgagaacac gcacatcggc 12540
ctgcgcccgt tccgcaacaa catctcgtac ttcggcgtcg atgccgacca gctcatgaag 12600
gagcggcccg acctcgcgcg ccagctcttc aacgagctga tgcagctctt cgagcagggc 12660
gtgctctcgc cgctgcccta ccgcgccttc gccgccaccg aggcggtgga agccttccgc 12720
tacatgcagc actcgcgcca gatcggcaag gtggtgctgt cgttcgccga cggcgtgaag 12780
gcgcagcccg cccggcccgc cgtgccgcgc gagctggcgc tgtcgccgaa cgccacgtac 12840
ctcgtcaccg gcggcctgtc gggcttcggc ctgcgcaccg cgcagtggct ggtggacaag 12900
ggtgcccgcc acctggtgct ggtctccaag agcggcgccg agtcgctcga gaccaaggcc 12960
gcggtggccg acctcgaggc ccagggcgcc acggtgatcg cggaggcgtg cgacatcacc 13020
gaccgcgcct ccgtgcaacg gctgctcgcc gaggtcgcgg ccgcgctgcc gccgctgcgc 13080
ggcgtgatcc acgcggcggc cgtgatccag gacggcttca tcaccaacat gacggccgcg 13140
cagatccgcg acgtgctggc gcccaaggtc ctcggcgccc gccacctgga cgagctcacg 13200
cgcggcgcca agctcgattt cttcgtgctg tactcgtcgg cgacgacgct gttcggcaac 13260
ccgggccagg ccaactacat cgcggccaac tgcttcctcg aggcactcgc caaggcgcgc 13320
cgcgtcgagg gccagccggc gctgtgcgtg ggctgggcg cgatcgggga cgtgggctac 13380
ctcgcgcgcc acgagaaggt gaaggaagcc ctgcagacgc acatgggcgg caccgcactc 13440
gagtcggagg ccgcgctcgc ggtgctggag cagctcctgc tcgccgacgc ctcgggcctc 13500
ggcgtgctcg acttcagctg gcgcaacctc cgccgcttcc tcccaccgc cgcctcgccg 13560
cgcttccgcg agctgggcgc gcgcggcgac gacgcgaagg acgacgacga ccggcgcggc 13620
gagctgcgcc gcctcgccaa cgagctggac gcctccgagc tctcggcgat gttcaccgac 13680
ctgctgcgcc gcgaggtggg cgagatcctg cgcatcccgc cggaccgcct cgatacgcgc 13740
cagcccttgc aggaaatggg catggattcg ctgatgggcg tggagctgct gaccgcggtg 13800
gaggcgcgct tcggcgtgaa cctgccggtg atgtcgctct ccgagcagcc gtcgatcgag 13860
aagctggtcg accgcatcgt gcgcgcgctg aaggacccga acgccggcac cgaggccgaa 13920
tcgcgcagcg accagatcga gcgcgtcgcc gcgcagtacg cgcccgagct cgactcgcgc 13980
caggtcgagg agctcacgca ggcggtcgat gacgcgcaag ccacgcagca gggacgtccg 14040
tgaaggcgcg caacctccgc ggcctctcgc aggcagtgaa ggagcagctg atccagcgcg 14100
tgctcgagca gcgcgtgcgc cgggtcgaga aggacaagcc ctcggcggtc gagtccgcca 14160
tcaacgcctt ccgccccgac cccgatgccc tgggccacat ctcggagcgc tactgccgct 14220
tcgacatgca cccggtctac cagcagatgc agctggtgaa ggaaggtgcc gcgaagctgg 14280
gcattcccga tccgtacttc cgcgcgcacg acggcgtggc cgcggatacc acgcgcattg 14340
ccgggcgcga gtacatcaac ttctccagct acaactacct cggcttctgc ggcgacccgg 14400
ccgtggatgc cgcggccaag caggccatcg acgaatacgg gaccctcggt ctcggccagc 14460
aggctcgtgt ccggcgagcg gccgctgcac cgcgagctgg agcgcgccat cgccgaggtc 14520
tatggcgtgc acgacgccgt ggccttcatc agcggccacg cgaccaacgt ctccaccatc 14580
ggccacctgt tcgggccgcg cgacctcatc gtgcacgacg cgttcgtgca caacagcatc 14640
ctgatgggca tccagctctc gggcgccaag cgcatggcct tcccgcacaa cgactggcgg 14700
gccctggacg agctgctgcg cgcgcagcgc cgccacttcg agcgcgtgct gatcgtgatc 14760
gagggcgtct acagcatgga cggggactac cccgagctgc ccgaattctc gcgcctgcgc 14820
cgccgccacc gcgcgttcct gatggtggat gaagcgcact ccttcggcgt gatgggcccg 14880
cgcggcttcg gcatccgcga ccacttcggc atggccggcg acgaggccga catctggatg 14940
```

```
ggcacgctct ccaagacgct cgcctcgtgc ggcggcttca tcgccggcga gaaggcgctg 15000
gtcgagcacc tgaagttcgc cgcccccggc ttcctctaca gcgtcggcat ggccccgccc 15060
gtggcggccg cggcgctcgc cgcgctgaac cgcatgcgcg aggagccggc gcgcgtggcc 15120
gccctgcagg cgcgcggacg cttcttcctc gaggccgcgc gcgccgcggg catcgacgtg 15180
ggcctctcgc agggcatggc cgtggtgccc gccatcaccg gcagctcgat ccgcgccggg 15240
cgcctggccg aggcgatgtt ccagcgcggc atcaacgtcc agccgatcgt ctacccggcc 15300
gtgcaggaga actcggcgcg cctgcgcttc ttcgtgagcg ccacgcacac cgaggagcag 15360
ctgcgcttca cggtccgcga gctcgccgac gcctggcgca agctctgagt ggcgggcccc 15420
cggctgcgcg tcgtcgtctg cacccacggc ggcctgaacg gcgcgctggt gctctcgcgg 15480
ctgctggccg cccccacgct cgaagtctcc gcgctggtga tctccagccg cgcgcgcggt 15540
gcccatgagt cgtatggccg tgccgcgctc ggatacgtgc gcgcgagcgg cgtcgcctat 15600
gcgctgtacc tctggtgcgc cacggcgctc gccgacctgc tgctgcgcgg cacgtccgag 15660
ggcccggtcg cgcgcatcgc cctcgcgcgc ggcatcccgc tcctcgccac gccgcgcgtg 15720
aacgatgcca ccgcgcgcgc cttcatcgcg ggggcggcgc cggacctgat cgtctccgcg 15780
ttcttcaacc agcacatcga cgccgacgtg gctgcgctcg cgcgcgtggc cgcggtcaac 15840
atccacccgt cgccgctgcc gcacttccgc ggcgtggatc cggtgtcctt cgcgcgcctg 15900
cgcggtgccg agcgccacgg cgtgagcgtg catcgcatcg aacccggctt cgataccggc 15960
gcgctgctcg cccaggaaac cgacgtcgag gccccggca gcgtctttgc cgccaccgcg 16020
gcgctctatg accgcgggggc ggcgctcctc gccggccgtg ccgcggcgct ggccgccgac 16080
ccgcgcggaa ccccccagcc ggcgggggggc tcctacgact cctggcccac ccgcgcccag 16140
gtcgccgcct tccggcgcgc cggcgggcgc ctgctccgcg cccgcgacct gtggcgcctc 16200
gcgcgccggg ggccggccgc tttcgtaata gaatcagcgc ggtagcggcg actcacgccg 16260
cgcttttccc cgaccccccg aggcccatga aacactggct gaagcaacac gcaatcttcg 16320
cgctcaccgt cctgctgccg accgtggccg cgatcctcta cttcggcctg atcgcctccg 16380
acgtctacat ctccgaatca cgcttcgtgg tgaggagccc ccagcgccag gtgcagaccg 16440
gcctggtggg cgccctgctc tcgggcaccg gcttctcgcg ctcccaggac gacacctact 16500
cggtgcacga cttcatcacc tcgcgcgacg cgctgggcga gctggacaag aagctcgccg 16560
tccgcaagct ctacacggcc gccaacatcg acttcatcaa ccgcttcccg gggctcgact 16620
gggacgacag cttcgaggcg ttccaccgct actaccagaa gaaggtcacg atcgacttcg 16680
acaccgcgtc ctcgatcacg gtgctgcgcg tgcgcgcctt cgagaaggcc gactcgcggc 16740
gcatcaacga cctgctgctg cagatggggcg agcgcctggt gaacgagctg aacgagcgca 16800
gccgccagga cctgatccgc ttcgcgcagg ccgaggtctc gctcgccgag gacaaggtga 16860
aggatgccgc gctggcgctc tccgccttcc gcagcaacca gtcggtgttc gagcccgacc 16920
gccaggcctc gatccagctg cagggcgtgg ccaagctcca ggaggagctc atcgccaccg 16980
agggggcagct ggcacagctg cgcaagctct cgcccgacaa cccgcagatc ggcgcgctcg 17040
agaacaagtc ggcggcgctg cgcgtggcga tggcgcgcga gtccgccaag gtgaccggcg 17100
gcagcggctc gttcagcgcg cgcgccccgg cgttcgagcg cctcaccctc gagaagggct 17160
tcgccgaccg ccagctgggc gttgccctca ccgcgctgga ccgcgcgcgc agcgaggcgc 17220
agaggaagca gctctacctc gagcgcatcg tgcagcccaa cctgcccgac gaatcggtcg 17280
agccgcgccg catccgctcg atcttcaccg tgttcgtgct gggcctggtg gcctggggcg 17340
tggtgagcct gctggtggcg agcgtgcgcg agcacgtcga ctaggccgtg gagcagaacc 17400
cttcccctcct gcgcgcctc ggggtgcagc ggcgcgtgct ctacgcgctg ctgatgcgcg 17460
aggtcatcac gcgcttcggg cgcgacgacc tcggcgtgct gtggctggtg tcgagccga 17520
tgatcttcac gctcggcgtg acggccctgt ggaccgcggc cggcatgaac cacggctcgt 17580
cgctgccgat cgtggccttc gcggtcaccg gatactccgc ggtgctggta tggcgcaact 17640
gcgcctcgcg cgcctcgatg gccatcgagg ccaacaaggg cctgctcttc caccgcaacg 17700
tgcgcgtgat cgacgtgttc gtgacgcgca tcatcctcga gatcgcgggc gccacggcgt 17760
cgttcaccgt gctgggcatc ttcttcatct ggatcggctg gatgccgatg ccgatcgaca 17820
```

```
tgctcaaggt cgcgttcggc tggttcatgc tgtcgtggtt cggcgccagc ctggcgctcg 17880
ccatcggcgc cggcaccgcg tactcgctgg tggtcgagcg gctgtggccg ccgacggcct 17940
acctgctgtt cccgctctcg ggcgccgcct tcatggtcga gtggctgccg gagaaattcc 18000
agcagttcgt gctcttgctg ccgatggtgc acggcaccga gctgctgcgc gacggctact 18060
tcggcaacgc ggtgcgcacg cactacgacg tgggctacat ggcaacggtc tcgctgggcc 18120
tgtcgctgct cggcatgcac ctggtgcgca tggcggccaa gcgcgtggag gccgcgtgat 18180
ccgggtcgag aacgtgacca gatgtactca cgcgcctcgg cccagaacgt gctgacgcgt 18240
cagcttcgag ctcggccgcg ggcgcaacat cggcatcctc ggcaagaacg cgccggcaa 18300
gtccacgctg atccgcctga tcagcggcgc cgagagcccc accaccgggc gcatcgtgcg 18360
cgagatgagc gtgtcctggc cgctcgcctt cggcggcgcg ttccagaccc acctcaccgg 18420
gctggacaac ctgaagttcg tctgccgcat ctacggcgtc gactaccgcg acaaggtcgc 18480
cttcgtcgag gacttcaccg agctcggcgt gtatttccgc gagccggtgt tccactactc 18540
gcacggcatg atcacgcgcc tcgccttcgc gctgtcgatg gcggtggagt tcgactgctt 18600
cctgatcgac gaggcgatgg tggtgggcga cgcgcgcttc cacgagcgct gccacgtcga 18660
gctcttccac aagcgcaagg accgcgcctt catcctcgtc acccacgacg ccaaggtgat 18720
caagctctat tgcgagagcg cctgcgtgct gcacgagggg cggctgctgc ccttccccac 18780
cgtggacgcc gcctacgagt tctacatgaa cgaggtcatg caggacctcg ccccggaggt 18840
tgcctgaacg cccccgcgcg gccgctgcgc gtgctcctcg agctgcggcc ggcgttcgac 18900
ggccacgccg gcatcccgca ggaggcacgc ctgctgttcc gcggcctgcg catgatcgag 18960
ggcatcgacg tggagggcct gctgcagcac agcggccacg tgctcgccaa gggcctgccg 19020
ccgcgcggcg gcggcgacct cgcaccggac cgccagctga accgcctgtc gcgcgtcgtg 19080
gtctccacca agcaggagct caccaacgcg cacgtggcga cggcgctgat ggccattcgc 19140
cagctgctgg gcgcgcgcga gaccctctac cgcttcgacg cggcgcactt ccgcgacttc 19200
gtctggcagg cgctcttcgc gcgcacgctg cacgccgcgg acttcgactc ggtcacccgc 19260
gccggctttc gcgtcgcgcg cgtgccgtgg acgggcatgc accgctgcgc gctggtcacg 19320
cgcaagctcg gctacacgct gcatccgcgc atcgacactt ccgatttcga cgtgctgatc 19380
gccgagacgc cctacccggc gcgggtcgcg gacggcacgc gcctcgtggt gcgctaccac 19440
gacgcgatcc cgctcctcat gccgcacacc atctccgaca tgtcctacca ccaggcctcg 19500
cactaccggg cgctgcgccg caacgtggcc tcgggcgcgc acttcgcgtg cgtctccgaa 19560
gccacgcgca aggacctgct ctcggtcttt cccgaggtcg aggcgcgctc cagcaccatc 19620
cacaacatgg tctcgcacca ctacttccgc gagaccgccg acgccgggcg catcgaggag 19680
atcctgcgaa cgcgcgcgag cgagcgcatc aagggcgcgc ccaacggagc cgccaccgcc 19740
gtcgcacgat cgccctcctt cctcgcgcgc gccgcgcaac ccagcgaccc cggctacctg 19800
ctcgccgtgg cgacgatcga gccgcgcaag aaccacgcgg gactgatcgc cgcgtgggag 19860
cacctgcgca cctcgcgctt cccgggcctg cgcctggtgg tggtcggcat gcccggctgg 19920
ggctacgagg cgctggtagc caagttcaag ccgtggctcg cgcgcggcca gctcttcgtg 19980
ctcgaggacg tgcccgcccc cgagctgcgc ctgctctact cgcacgcgcg tgccacggtg 20040
tgcccgagct cggcgagggg cttcgatttc tccggcgtgg aggcgatgcg ctgcggcagc 20100
ccggtgatcg cctcggagat cgccgcgcac cgcgaggtgt accgcgacgc cgccgagtac 20160
tgcagccct attccgtggc cgacctcgcc gaggcgatcg ggcgcgtcat cgacccggcg 20220
gcgacgggcc tgcgccaggc gctggtcacg cgcggcaccg aggtctcgca gcggtacacg 20280
cccgaagcca tcctgccgca gtggcgcgag tacctgctcg gcacggtgcg cgcggaggcc 20340
ccgtgagcga tccgaaggcc accgaccagg cgtgggagga gtggggccag cgcgacccct 20400
acttcggcgt gatcaccaac ccgcgcttcc gccgcgggca gatggacgag gaagcgaagc 20460
gcgaattcct cgcgtccggc cgcgtccatg ccgactacgt gatgcgcatg gtgcacgcgc 20520
acatcgcgcc cgacttccgt ccgcgcacca tcctcgattt cggctgcggc gtgggccgcc 20580
tcgtcatccc cttcgccgcc caggccgagc aggtgaccgg cgccgacgtc tcccctcga 20640
tgctggccga ggctgcgcgc aactgcgccg·agcagggcgt ggccaatgcg cgcctcgtgg 20700
```

```
tctcggacga ctcgctcacc gggctgccgg gccccttcga cctcgtgcac tcgttcatcg 20760
tgttccagca catcgatccg gcgcgcgggc gcgacatctt ccgccgcctg ctgggcaccc 20820
tggcgagcgg cggcgtgggc gcgttgcact tcgtctatgc gaagcgtatc tacgcggcga 20880
cgtacggcgt ggcgccaccg cccgagccgc cgccgccccc gccgccgccg cagcctccac 20940
ccagccgggc ggagatcaag gccgcgagca aggcgcgggc cgcgctggcc acgcaggcac 21000
gggcgccgga accggcaccc aatcccgacc ccgacatgca gatgaattcc tacccggcgg 21060
cggagatgct gttcctcgtg caggaagccg gcgtcacgcg cttccacgtc gagttcaccg 21120
accacggcgg cgagctgggt ctcttcctct tcttccgcaa gccctaggcg tccttcgagg 21180
cgaaggagag gcccagcgcg gtgatcgtcg cggcgtgctg cggggcgacc gccgccgcgg 21240
ccgcgacttc ggcagccacg cgctcgaggt cgccgcgcga cccgtacagg cgcacgcgcc 21300
agtagtgcac ccagtaggag tccggacagg ccgacgcgat tcccagcgcg cgctcgacct 21360
cggggcccgac gcgcccctgg tgcacttggt gcgaggccca ttgcaggttc acggagggct 21420
cgaggggcag gacggcgagc agcgcctcgc tcaccgccag cgcgcgcgcc tcatcgccgt 21480
cgtcggcgag gcgcatcagc gtgggcaggg actccggcgc aagcgcgtgc gcggccgcga 21540
gcgaatcgcg cagcatcgca gcgtcgccgc gcgcgagcgc caggcgggcg cggtggtagt 21600
gggtccagaa cgaaggcgcc gcggcatcga gcgcgcgcgc gacctcggcg cttgcgtcgc 21660
ccccggagag ccgggacgcg gcccacaaca ggttcgcctc cgccgcgtgc gccggcacgc 21720
tccgcaacgg cgccagcacc gcggtggcga cctcgcgcga accctcgcgc gacgagaagg 21780
ccgagcgcgc gaactgcatc agcgattcgc cgagcgacga atgttcggcg cagagccgcg 21840
cgggatccac cgcgccgagc gcgcgggcgg cacgcagcgc cgcctcctcg tcaccgcgaa 21900
gcgcgtgcag ccgcaggcga tggaacgcgg cccactccgc cgacatcgac gccgccccgc 21960
gctccagcgc accctcgacg gcgccatcgg tgcgcccgcg cgcgatgtgc aggatggccc 22020
aggtcgcctg cgcgccgggg aggttcgtgc cttgcgcgac gagcgggtcg aggaccgcga 22080
gcgccgcggc ggcgtcgcgc gcgccactga agtcgttggc caggccgatc agcgccccgg 22140
gcgccgcggc ggcgtcgcgc gcgccactga agtcgttggc caggccgatc agcgccccgg 22200
cgagctcggg ggccgcgcgg aaggcggccg ccgcgtcgcg ccccaccgcc gcgaggtcgc 22200
cgcgcagccg gtagaggcgc gcgcggtggt aatagcccca gaacgggtcg gggctcgcga 22260
tcagtgaatc gaacgagcgc tcgacctcgc cgcccacgtt gccgaggtcc acctggtgcg 22320
cgccccacag ggtcgccgca tgcgcggcca tcggctcgtc ggggacgaag gcgcggtgga 22380
tcagctcgac gtcgccgaag cgccgctcgc gcaacgcgag ctccagggcg ttcagccgct 22440
cgaggctcgc cttgtactcg cgataggcat cgcggtcgcg gtcgtattcc tgcgcggcac 22500
tgtcgtccat cagcgaggcc gtgagccgct ggtcgccgtg aacggtgtcg tatgcgtagg 22560
ccgcgcgatc gtcgtagggc agcgtggccg cgcgcagcga cccgaacgcg gcgcggtcgc 22620
ggaatcccag cttgcccagc acgcgggggtg ccaggtcgtg ggcttccatc tcgcccgggc 22680
tccagtcgcg cagcgccggc acgttcgagt agagccacaa cggcgtgcgg tagagctgcg 22740
ccaggtccgc gtccgggcgg ccctgcgcct cgaagatcgc cttccactcc gcgagtccct 22800
cgtggccggc ggagaagatc ggcacgtggt cgccgaagat cgccagcacg tagggccgcc 22860
cgcgcacctc gagacgctcg accagcgagc gcagcgcgcg gtcggtttcg gccatcaccg 22920
cgcagtactt gtcgatcgcg ccctggcgca cgtgcccggc gaggaagtgc ggcgccacgt 22980
agcgcggctc ggcgcccttg tcgcccaggt agggtccgtg gcccatcagc gagacgccga 23040
acacgaagcg cggccgcgaa tcgtcgagct ccgcgatcac ggcatccatc acggccgaat 23100
ccgcgatgcg ctcgaactgg cgcggcgcat gggcgaacgc ctcgtcggcg tggaacgccg 23160
cgaagccgag cttcttgtac acgcggtcgc ggttgtagaa ccagcggtgg aacggatgga 23220
tggccacgct gcggtagccg tggcggcgca gcatcgcgcg gcaccgcatac acccggtcgc 23280
gcaactgccc catgtagggc agcgcgcggc agccttcctc gcccaatgcg atgccggtga 23340
ggatctcgaa ctcggtgtcg caggtgttgc cgccgatcac cttcacgcgc aggcggccgg 23400
tggtgcggcc cggccgcgtc gcgccgagga acggcagcgg gtcgcgcgcg agcggcacgc 23460
cgaacgggcg caggtcggcg agcgactcca tcatcaggaa gatgatgtcg gggcgctccg 23520
cggtggccgg gccgctcgcc gagcgctcct ggcgctccag cagccgctcc acgccctcgc 23580
```

```
gcagccgccg cggcgcgacc gggcgggcga tgctcccggc gaccagcacg gcgtgcagga 23640
cgaagcccag gcgccgcacg ttcacgttgt ggtcggtggg cgaatacgtc atcgcgtcga 23700
actccaccgg ccagcggcgc aagcgtccgg cgacgaaggc gacgccgagc gcggcgagca 23760
cgagcagcca ctgccggcga tcgtcgtacg cgtcgtgcag gtcgatggcc tcggtgatct 23820
cgaacgcggc ccacgccagc gccgcgagca ctgcggcagc ggcaagcgcg cgcaccgcca 23880
cgcgcgggcc atgcaggtac gccggcgcga acaccagcgc ctcgcgagcg ctcaggaaat 23940
cgcccgggcc gaccacgcgg ccgaagaacg cgagcaggcc tgcgttcacc gcgtagagcg 24000
cgcccgccag cgccacggtc cccaccgtcg cgagcgcccc gaccgcgacg aagagcagcg 24060
catggaagat ggcgagcagg ccggcgttcc aggccaggcg ccgcgcgcgc gccacgggtt 24120
ccacgtcata gcccaccacg aggcgccgct cgagggcggc ctgcagcacc gcgtaggcga 24180
gcacgcagac cagcgtgccg cccagcagca cgccgaggaa acccgccaaa ccagccgtgt 24240
cgtcgaggat catgggacca ccaggacgt cttccggcca gtctataacg ccccgcccca 24300
tgaaaaacgg cgaggcccgg gggcctcgcc gtggagggaa ccgcgcaacc tcagggttgc 24360
gggggtgcga ccgaccagcc gccggggccg atcaggtttg cgctgccgat ggtggcgagc 24420
ccgttcatca ggcgcaccgt gacgctgccg tcggtgtggc ggaagaccag gtccttgcgg 24480
ctgtcgccgt tcaggtccag cagctgcgtc acgttccaac ccgtgcccgc cggcagcgca 24540
tcgcccgcgg tgatgatgga aacaccatcc atcaggcgga tgtgcacgcg gccgtcggtg 24600
tggcggaaga ccaggtcgtc cttcgcgtcg ccgttcaggt cgcccaccag gttcaccgac 24660
cacccgctgc cggccggcag gagctgggcg ctggcgccga aggtggtgcc gttcatcagg 24720
aacaggtgcg cgcggccgtc ggtgtggcgg aagatcatgt ccgcgcgtcc gtcgccgttg 24780
aggtcgccca ggtggcttac cgtccagccg ctgcggccg agaggaagcc cgctccggcc 24840
gtcaccgtgg tgccgtccat gatgtagatg tagccgcggc cgtcgctgtg catgaagacc 24900
aggtccgcct tgccgtcgcc gttcatgtcg cccgcctggg tgagcgtcca gccggtcccc 24960
gccccgaaga gctgggcgct gccgatgatc gtcgtgccgt ccatcaacca gatgtgcgcg 25020
cgtccgtcgg tgtggcgcag gatgatgtcc gccttgccgt cgccgttcag gtccgccgtg 25080
tggctcacga tccagccgag cccggcgggc agcagctcct tgccggccgt caccgtagtc 25140
ccgttcatga tgtacgcgta ggcgcggcca tcggtgtggc ggaacaggat gtcggcgcgc 25200
aggtcgccat tgaggtcggc cgtctggctc accgtccagc ccgcgccggc gccgatcagg 25260
ttggccgagc cggtgatggc ggtgccgttc atcgtccaca ccgcgatgcg cccgtcgctg 25320
ttctgcagca cgaggtcgct cctgccgtcg ccgctcaggt ccgagacgac cggcttggtg 25380
ggcggcgtga cgccaccgga aaccggcagt gccagcccgt cgatccaggc gcggtcggcg 25440
cagttggcgc catccggggc cggcgggttc caggtcggcg agttcaagg cgtggagagg 25500
aagttcgaga aacgccacac cagcgtgtgc gcgcccgcgg tcaccgggaa cgacaccagc 25560
ttccagcccg aggccgtggt gcccgcgtcg ctgaacacca ccgtgccgtc gatcaggaac 25620
tcgaacttgc cggcgttggg gaagctcgac acgcgatagg cgaacgcgac gttgcccgcg 25680
agcagcgtgc ccgcgaacga gaggtcggag ttcaccgtcg tcgagttggt cgggtcgctc 25740
gtcaccacct gcgccgaacg aaggctggtc gcgccctcga aggcctggtc ggagccgacc 25800
gtccatgcgt tggcgccgcc cgaggtcgtg aagccggcag gcagcgtgcc acccgtgggc 25860
cacggcgaga ggatgatgcc ggtcgcggtc gccgggccgc cgatgaagac acccgtggaa 25920
cccgtcgcat tggagagcgc gacgttgaac gtctcgttgc cttcaccgat accgtcgctc 25980
gccaccggga ccgtgatggt cttgggcgtg gtctcgccgt tggcccagct gagcgagccc 26040
gaggtcgcgg tgtagtcggc gcccgaagtg gcggtgccgt tggaggtggc gtagctgacc 26100
gagatggcac ccgccgagcc gccgatgcgg ctcacggtga gcgtcacgtt gccggcggtt 26160
tccgcggccg cgaaggtggt gccggtgaac tgcaccgaac cgggaacggc gggcggggtg 26220
accgtggact tgagcgccga gagtgccgcg cggttgttgt tcagcgagag cgcgtcgttg 26280
gcctgcgcgg tgccgcacgg ttgcggcgca ccgacacccg tcgtgcacga gaggcccggg 26340
ttggagaagc ggtagacctt ggtcgtcgag tcatgcagca gcgacatgat ggtggtgaag 26400
ttcgccgcgt cgttcgtcgt gcactcgggg cgctggctca cgccgcaggc gccgccgccg 26460
```

```
gagccgttgg gcagggccga gttgcaggtg aggttgccgc tcgcgcaata gaagtggccg 26520
tacgagtacg cgaacgcacc ttgcggcggg ctcgcggtgc cgcccgcttc ccaggcctgc 26580
gaggggcggt cgtggcggtt gcccatggca tggccgagct cgtggatgaa gacccactcg 26640
caaccgaaca ggcagccgga agcaacggcg taggcaaagt tggggtccgg cgtgctggtg 26700
ccgatccagg ctgcgccgtt gccgccgaag tcggcgccgt tgcgcatgaa ggccaccatg 26760
tccgcgccgt attgcgtgcg gatggcctcg atgttgccga acgtcgcggc gtcgaagctc 26820
gcgtcgcccg gggtgatggc atgcatcgcc gtggtgtcgt cgatggcgtc gtcgtagttc 26880
acctgcgtcg cattcaccag cgcgcagcgtg atggcgacct cgctgtcggc atacgcagtg 26940
ttggcgcgcg tgacgaggaa gttcaggcgc gtcatcaggc tgccgccgag gcgctggccg 27000
aagccgcgcg agtacacgat catcaggtcg atgacgttct gcggcgtggg cgtcacctgg 27060
gcgagccgac ggtattcacg ccggggatgg cgatcagggt ctcggggctg ccggtcgcga 27120
cgacgcgcga cttcgagccg gtcgtggcca cggcgtggtc ggcgccgagg ctgatgatcg 27180
gcatgctggc ctgctcggcc gtcatgtcga ccaggtagtc gacggtgccg cccggaatca 27240
ggcggtactc accctgcggg gtcgagaaca cgccgaagga gccttcaggc ccggtcgtga 27300
cgatcgcgcg gtggttgatg ccgtcggtct tgctcttcgc gatccacgac gtgatgccgt 27360
cgccgtgggc ctggagcaac tcgaagacat acgagtggcg caccccgttg gggagcgaca 27420
gctcgacctc ggagcgcgga ctcagggcat ggagggccgc ggcgttgaag cgcaccgctt 27480
gccgggcaac ggcgctgccc ggcgccggca tcgcggcggc gggtgcggag accaggatct 27540
ggggaaccgc ggcgaaggcc acggaagaga agacgatcgc cagcgcgccg aggaaggcgg 27600
ctgtgatccg ttgcgtgaat gcgttcatgt ggagctccgg aagttgaccc atgcccaatc 27660
cgctatgtcg cggagatgtg dacaaaaggt atcgagcggg cgtgacgacc cgccccgga 27720
gggatgctcc aaaaggacta cgagggtgct acggctgggt tggtggcgaa ggccgaagga 27780
gaactccttg tggtctccgc gcgacttaag gttgcgaggg aaccaccgac cacccaccgg 27840
ggccgatcaa gtccgcgctg ccgatcgtgg tgacgccgtt catcaggcga atcgtcacgc 27900
gcccatcgtt gtggtggaag accacgtcca tgcgcccgtc gccgttgaag tcgaggagct 27960
gcgtcaccgt ccagcccgcg aacggcggca ggatgtcggc ggcgccgagg atcgcggtgc 28020
cgttcatcag gcgcacgtgc gcgcggccgt cggtgtgcgc gaacacgagg tccgcccggc 28080
cgtcgccgtt gagatctccc accaggttca ccgaccatcc ggtgcccgcg ggcaacagct 28140
cggcgctggc gccgaacgtc gtgccatcca tcaggaacag gtgtgcgcgg ccgtcgttgt 28200
ggcggaagac gaggtcggcc ttctggtcgc cgctgaagtc cgcgacgtgg ctcaccgccc 28260
agccgctgcc gggcgcgagg aagcccacgc cggcggtgat cgccgtgccg ttcatgatgt 28320
atgcgtagcc cgcgcccatcg acgttggaga agacgatgtc ggccttgccg tcgccgttca 28380
tgtcgccggt gccgaccacg ttccatcccg agccgcgggg cagcaggctc gcactgccgc 28440
tgatcgcggt gccattcatc agccagatgt gggcgcgtcc gtccgtgtgc tgagcagcag 28500
gtcggccttg ccgtcgccgt tcaggtccgc cgtgtggctg atggtccagc ccgtgccggc 28560
ggggaacagc tccttgccgc ccacgaccgt gagcccgttc atctggtacg tgtagatgcg 28620
cccgtcggtg tgctggaagt agatgtcggc ctggctgtcg ccgttggaat ccgcggccat 28680
tgatcacgct ccagccggcg ccggcgggaa tcaggttcgc cgaaccagtg atcgtcgttc 28740
cgttcatggt ccacgcggcg atgcgcccat cggtgtgctg gaacaggaag tcgctctgcc 28800
cgtcgccgtt caggtcggtc gccgcatggg tcgggaacgc caccttcgcg aggaaggcat 28860
cggagaaacc cgtgagcaac gtcttgtacg cgccggcgt ggtcggatag ttcgccgacg 28920
acgtgcggcc ggccacatag gccgcgcccc gcgcatcgac gccgacgccg taccccacct 28980
cgacgttgtt gccgccgatc agcgtcgaga acacgatgtc gccgttgcgc acgcgagtga 29040
ggaacgcgtc ggcatcgccg ccggggccct gcaccgggtt cagggtcggg aagccggcga 29100
cctgcgtgta gcccacgacg agcgcatcgc cattcggcgc cagggcgacg tccaggggat 29160
tgtcgtaggc cgagccgccg aagatgcccg acgattcgag cgcgccggtg ccgagaaagc 29220
gcgtcacgaa cgcatcgacg gtgccggcga aattgcgcac cgttccggtc tgcgggaagc 29280
tgggcaggcc ggtatcgccg acgacggtcg cctggcccgt gaccgggttc accgcgatcg 29340
```

```
cggtcacgtt gtcgccgctg ccgccgagga acgtggaata cgcgaccggg ccgccgctcg 29400
caccgatctt ggtgacgaag ccgccaccgg gttgggtctg gtacgcattc acgttgcgga 29460
acgtcgagga gcgcgtggag cccgcgacgt acgtgttgcc cgcattgtcc gttgcgatac 29520
cgcggccgat gtcatcaccg ctgccgccga gatacgtgga atagacgaag ccgccgttga 29580
acgagagctg gtgacgaacg cgtcgcgatt gaggccggtg ccgacgaaca cggctgcagc 29640
gcgctgccgc cggtcgggaa gccgcgcgcg gccgaacccg tgacatgcat gaagccgttg 29700
ttgtccacgg cgatcgccat gccctcgtcg tcgccggagc cgcccagcag ccgcgagaag 29760
atgacgttgc cagtggtatt cagcgccgtg acgaatgcat cgcgcccgcc ggcatggttg 29820
gtgcaatcgg tgacggcgaa gcaatgcggg aacgcgttgg tgccggtgga accggtgatc 29880
gcgacgaact cgttgccgcc ggaggcgccg agcgccagcc cgcggagata cgcttcggcc 29940
ccgagcgtgg tcgagtacac cacctgctgc gtggtggggt tgaacttcac caccgtggct 30000
tcgtacgtgc ctgcggtata ggtcgcgagc ccgacgtaca cgttgccctg cgcgtccacc 30060
ttgaccggcg tggtccactg ctcgtcgccg cgccgcccca ggtacgtgga gaacgccatc 30120
accgggtcga tcaccagcgt gcgggtgctg tcgtagtcgg cgatcacgaa gccggcctcg 30180
gcgccctgct cgcccacgaa gagctcgaag cgcgcggcca ccggcacgcg cgaatcgccg 30240
acctgctgga aggctaccgg cgcgtgctgc gtaaattcct cctggcccac gcgcatgcgc 30300
aggttgccct cgccatcgat ccacgcgtgg tcggcggacg agaggtcgag gcggatctgg 30360
cgcggatcgg cgcgcggggc gaccacgaag tcgtattcga gcgtgccctc cttgccgtac 30420
acggtgaggt ccacgccgcg atacaggtcc ttcagcgaca cgcggccgaa gtgcggcacg 30480
ttttcgcgcc ggccggcggc cgtggtgccg ctgtaatagt ggctgagggt ggcccggggc 30540
tcctcggcct cgatgaccgg atcggaagcc cccgcgaagc gcacgcgcaa cagcgaggcg 30600
ctgcagctt tcctggagcg tggctcgaag gcaatgccat cgcggctcac ggagacgcgt 30660
ccggcctgcc cgcgcgagac gtagagcgca tccgggccga actggccctc gttgcgctcg 30720
aacgtgatgg cgcgcgttcct gagggcatca gggacgacgg catggacggc ggcgggagtg 30780
gcgagggcgg ccgcgacgag cacggccagg gggcggagcg ccgggcgggc gccgctggag 30840
ctgcgggtca tggggcagtg accttctgt tgttatggct gcttgttgtc tttgcgggat 30900
ccacagctcg gatccagctg gcgatactac cagcgagcga gtggttttc gtgacccagt 30960
cccgcagtcg ggccccggcg acaacggcgg catcgcggtc ctttgccagg gcacgcacgg 31020
cctcgatcca gtcgccgtcg cgggccagca cggcggggggc atcgcggtag ggctcgagat 31080
ccgaggccac caccgggatg ccgaggcagc cgtactcgag cagcttcagg ttgctcttcg 31140
cgcgattgaa cgggttgtcg cgcagcggtg ccaccgcgac gtcgagtgcc agcgacgcga 31200
gcttctccgg gtactgcgcg atgggcacca tgtcgtgcac ttccgcggcg aacggcgcca 31260
gctcgggcgt gcacaggccg aggaacaccc agtcgatctc gcgatgcgtg gcgcgcacca 31320
cgggctcgag caacttcagg tcttccccat gctgtttcgc ccccgcccag cccacgcgcg 31380
gacgcgcgcc acccgcggga cggttcgcga ggccggccca acgctccgca tcgatcgcgt 31440
tcgggatcac gcgcacgtcc ttcgcgccgc ggccgaaggc ctccgcgagc ggtgcggtgg 31500
acaccacgag gcggtcgcac aacgccaccg cgcgcgcgat gcgctgggcg atgtccgggt 31560
agatcgtcgc cgcatacgga ttgccgggcg gcagttgcgt gagcaggtca tccaggccca 31620
ggaccttcag cgcgcgtccg tggcgcgcga gcacctcgag cgaggtgagc tggtagttgt 31680
ggaagaagtt gtgcgcgagc acggcgtcgg cgtcgaggcg ctgccactcg acgcggttgg 31740
gcgcgcagcc cgtggcgtgc tcgcccatca tcacgacctc gacgcggccc gcgcgctcca 31800
gcgccgcgca cggctggcgc acgcgcacct cgccggagcc ccagcggtcg aacggaaacg 31860
cgcagagccg aaccgcgtgc gtgtcgttgc cgcgcggcgc gaagcgttcc acgggcccgt 31920
agccctcgcc gcccatgcgc aacgccgggt ggtagtgcgg atcgtcgtcg agcaccggct 31980
gccagcgctc gcgcatccac gcctcttcgg aaggaatgac gagcttcgcc accggaacga 32040
cggcatcggc gcgcgcatcg gcgagcaacg ggaagtccgc cgcgatgccg gggcgcgcaa 32100
ggatgtcgaa gcccgctgcg cgcagcccga ggcacaggtg cgccatcgcg aacggccccg 32160
cgcgctcgat ctcgtgcagc gcctgcgcgg agagcgccgc gtcgcggttg accagcgcga 32220
```

```
gacggcccgc cgcgacgctc acctcgcgcg gcgatgcgta gagtgccgcg aggcgctcct 32280
cgccctgcgc gcgcggcgga gcgcccgaga cggcccacgg cccgccgccc gcgatttccc 32340
agccgggaac gcgcacaccg ccggggccca gcaggtcggg cgcgagggca cccgtggagg 32400
ggcctgcgat gccctgctcg aggcgctcga gccagccggc gcggaattcg ctgcaacgcg 32460
catcgacgat cgcgatccac agcgtcgccg cgttgcggat ggcgcgcgcg aacgcggagc 32520
cggtttcgcc gcccaccgga accaggtgcg tgcgcacgcc gcgcccgccc agctcaccgg 32580
gccagcgttt cggcgccctt cgcgtcgaga tcgacgtaca cgtggccgat acgttcgcgg 32640
cagcccttca cgagcgcctc gatgcacgcc accagccgct ccggcgcgtg cgtcgcgtca 32700
atgaaggcgc aaatcgtgcc ggccccgatc gcgcgcgccc acgatgcggg cgcatcggac 32760
gcagctggca ggaacgcggc ctgcgtgcgc gtgcggcgga agtgctccgc gaccacgcgc 32820
tggtggcgct cgcgcgcggc gtcgctttcc tgcgcggcga tgttcaggtc gaggcgatgc 32880
acgagcagct cgcgcagtgc caccacggcg ctcgcacggc cctcctccgc gaggcgcagc 32940
gcgagatcgg tgacgcccgc ccagccgccg gcgcgcagcc cgccacgctc cagcaacgca 33000
gcgcggcgca gcaccgcgag cccggcgagc ggctggtgtc gcgcgaactc cgcgccgaac 33060
gtgggggacga acgccggatc gtagcgttcg ccgtgcgcgc tcagggcatc actgcagccg 33120
tagagaagct cggcgttcgg gttgcgtgcg accgcacccg ccaccgcgct cggcgcgcca 33180
tcggccagcc ggtcgccctc ctccaccagc acgacccatt cgcccttcgc gtccgcgatc 33240
gcgcggttga cgcgcccgag cgccgcgtcg ctcgcgtcgc gctcgaggac cacggtctcc 33300
agggcgatcg cggcagtgct cctggccggc aaggcctcgc gcgccatcca gtggtcgtag 33360
gcatcgccgc ccagcggcgt gcgctcgcgg cggcgtggaa agccgaggca cgccgcggcg 33420
aacaacggat cgatcgttgc gggcacagca tccaccgcgc catcgcgctc cgccgccagc 33480
tcgcgataga gcgccaggta ctcggccgca ttcgcctcga tcgtcttcgg gcggtggtac 33540
ttcaggttct ccgcgacgcg atcgagcgcc gtgcggtcgt cgcgcaggcg cagcaggagg 33600
gccacgatcg cgccgggatc gttgtgcggc agcaggaagc cggtctcgcc atcgcgcacg 33660
cgctcgggaa tggcgccatc gcgcgtgccc gccaccggga tgccgaggcg ctgtgcctcg 33720
ctcaacacca ggcagaaggt ctcctccatc gtcgagggca gcaggaccag gtcgatgccg 33780
cgaaggatgc gcggcaggtc cgtcgcggca taggcgccgc gctgcatgat gcccgccgca 33840
ttcgccgccg cttcgagcgc gagatcgcgc ggaccccagg cctcgatcgc gatgcgctgc 33900
cccgccagcc ggcgcgcggc ctcgatgatc aggtgcgcgc ccttgcgccc gccgaagcgg 33960
ccgaggaagg ccacgcgcag caccggccgc gcttcggggc gcggaagcag cggcaggtcg 34020
gcaatgccgt gcccgatgac gcgcagcttc gcttcgcagc tcgcgccgaa gcccttgcgc 34080
atgcgcatcg ccgcgtactc ggacgggcag aggatcacgt cggccgcgtc cacggcatcg 34140
gacgccgcgg cgaagcgcga ttcgaggtag ccggtcacgg acgccggcgc cgtcggcccc 34200
acgatcgcct cgctgcggcc gcgcaggcaa tccacgcatc cgccatcggc ccgcgcgcg 34260
gcggcgcggc cgcacggcag gtcctgcggc ccggtcatca tgttgtagtc ggcgcacagg 34320
tacgagaagt cgtgcgccgt gatgaccacg cgtgcccccg actcacgcgc gatgcgcggc 34380
aggcgcaggc tgttccagcc caccaacgac tggaagtgca cgacgtcgta gcccccggac 34440
agcagttcgc ggaaggccgc ctccacgccc gcgtcggcaa ccgccgcctc gtgcccgatc 34500
acgcggatgc cgggcgtcat gcgctcgggg ttcaggcgcg cgagtcgcac gtgcgtcgcc 34560
aggcgctcgg tggtgaagtc ggtccacgcg cgcgggccct cgggcagcac cagcgtcgag 34620
gcgacgtcct tgcgcagcag gttcaccagc gcgcgcgtgt gctcctcaat gccgccgcgc 34680
gagtcgaggc gatgcagcac ctgcagcacg cgcgggcggc ccggcatccg ctcgcgctcg 34740
gccgcggcgt tgacgcgctc ggcggcgcgc cgcagcgggt tcgcgatcgc ccacgcgcgc 34800
acgcccgcga cgtaggtggg ccagcggcgc tcgaggcgca cgcggttggc ctcgcgcgcg 34860
gcgtcgatgc cggccacccc gccgaacgag gcctcgcccg cgtggtgcac gtagacgtcg 34920
tcgcagcacg cgatctgcca gtcccggtcc agcgcgcgca tcgagaggtc gttttcctcg 34980
ccgtagccgc ggccatagag cgagtcgaac gggccgcatt gcacccacac ctcgcggcgc 35040
accagcatgc agaagccgac cgccgtcggc aggcgcggat actcgcggcg gctggcgcgc 35100
```

```
gcgacgcagg cggagatggc atcgacgtcc ggctcgccgt cgggccgctc gaagatatcg 35160
gcgacgccgg gcaggctgag gatcgtcgcg cggttggaga gcgggcacac gatgccgacc 35220
cgcgcatcgc tgtcgcggca acgcagcagg ccctcgagcc aaccggggcc gacctgcgta 35280
tccgaattga gcaccacgaa gtcgccgccg ctgccggcca ggcccgcagc gaccgcgcct 35340
gcgaagccga ggttcttcgc attggtgacg acacgagtgc cggcacgacg cgccgcgaag 35400
ctgcccagca acccggcgag gcgcggatcg gtgctcgcgt catcgacgag gacgatgccg 35460
tgcgcggcgc ccgtgtggcg ggccagcgac gcgaggcagc gttcggtgtc gtcgtggccg 35520
ttgaagaccg ggacgacgac tagcgccggc gccatccgac cggcgcttcg agctgggcca 35580
atcgcgccag caggcgctgg tgctcctgcg agagtgcctc gagctgcttc tgcgaggcga 35640
cgtaggcgcc tccatgcgct cccaggtgcg tgcgcgcctc ctcgagctgc cgcagcgact 35700
gctcgtggcg accggagacg tcgtggtcgc ggtcggagag cagcgagaag cgcccggccc 35760
ccgccgtcac cgcctcgcgc tcgccgcaca gggcgaggaa gtacatcgca tcggcgacgc 35820
ccgccgcggg tttcgcagcc tccggcgccg aagcttgcag caactgcacg gcgtcgggca 35880
ccgcgtgcag cggccagatc gccgaatacg catcgacgcg ttgcccgaac agggccacct 35940
ccgggaaccg cgccttcagc acggcgagga attcgtgctc gtagagctcg cgcacgtgat 36000
gcgggttgcg gtagtcgcgc tggtccgagt acacctcgcg gttcggcgtc gagaccagca 36060
gcaggccgcc gggcgccagc acgcgcttcg cttcgtcgag caggcgctcg ggatcgggaa 36120
tgtgctcgag cgtctcgaag gagacgagga gatcgatgct tgcgtcatcg cacggcagcg 36180
actcgcagcg gccctcgacg tactgcaggt tctgcgcggc accatagcgg cggcgcgcct 36240
gcgcgatggt ctcggcggcg acatccgcgc ccaccacgga cttcgcgcgc gtcgccagca 36300
gcgccgagcc gtagccctcc ccgcaggcga tgtcgaggac gcggcagcct ccggcgagcg 36360
gcaacgcaaa gtggtagcgg tgccagtgct cgtaccagat ctcgccctcg aagccgggct 36420
ggaaacgttc gtgttccatc gggagagtat aggtggatgt gcaaccgccc tccggagagg 36480
gcggttgcga tgatcttatt cgagatcggc tgctacggaa gcaccggcaa cggctgcggc 36540
ggcgccaccg accatccggt gcccgcgccc agcaggttcg cgctgccgag cgtggtgagg 36600
ccgtccatca ggcgcaccgt gatgcggcca tcgacattct tgaacaccat gtcgagcttg 36660
ccgtcgccat cgaagtcgag cagttgcgtg acggtccagc ccgcccccgc cggcagcacg 36720
tccgccgcgc tgaggatggc ggtgccgttc atcaggcgca cgtgcgcacg gccgtcggtg 36780
tggcggaaca cgatgtcacc gcgcccgtcg cggttcatgt cgcccacgtg gctcacgatc 36840
cagccggtgc ccgcgccgag gagctcggaa ccggcgccga acgcggtgcc gttcatgatg 36900
aagagatgcg cgcggccatc ggtgtggcgg aagaacatgt ccgccttgcc gtcgccgctc 36960
acgtcgccca ggtgcgtcac cgtccagccg ctggcggggcg agaggaagcc cgcgccggcg 37020
gtgatggtgg tgccgttcat caggtagatg tagccgcggc cgtcagcgtg gatgaagacg 37080
atgtcgtccc tgccgtcgcc gttgaggtcc ccggtggcca cgactctcca gcccgtcgcc 37140
ggccccagca gctgctggct gccgatgatg gccgtgccat ccatcagcca caggtgcgag 37200
cggccgtcgg cgttgcgcag caggagatcc gccttgccgt cgccgttcat gtcggcggtg 37260
cggtcgatgc tccagccgag gccggcgccc agcagctcct tgccgccggt caccgtgagg 37320
ccgttcatcg tgtacacgta cacgcgcccg tcggtgtggt gggatcctct agagtcgacc 37380
tgcaggcatg caagcttgag tattctatag tctcacctaa atagcttggc gtaatcatgg 37440
tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa ttcacacaac atacgagccg 37500
gaagcat                                                               37507
```

<210> 3
<211> 26
<212> DNA
<213> Artificial Sequence

●

```
<220>
<223> Description of Artificial Sequence: Sense primer.

<400> 3                                                                    26
ggsccskcss tsdcsrtsga yacsgc


<210> 4
<211> 23
<212> DNA
<213> Artificial Sequence


<220>
<223> Description of Artificial Sequence: Antisense
      primer.

<400> 4                                                                    23
gcbbssryyt cdatsggrtc scc


<210> 5
<211> 773
<212> DNA
<213> Artificial Sequence


<220>
<223> Description of Artificial Sequence: OriT derived
      from plasmid RP4.


<400> 5
gatctgtgat gtacttcacc agctccgcga agtcgctctt cttgattgga gcgcatgggg  60
acgtgcttgg caatcacgcg cacccccggg ccgtttttagc ggctaaaaaa gtcatggctc 120
tgccctcggg cggaccacgc ccatcatgac cttgccaagc tcgtcctgct tctcttcgat 180
cttcgccagc agggcgagga tcgtggcatc accgaaccgc gccgtgcgcg ggtcgtcggt 240
gagccagagt ttcagcaggc cgcccaggcg gcccaggtcg ccattgatgc gggccagctc 300
gcggacgtgc tcatagtcca cgacgcccgt gattttgtag ccctggccga cggccagcag 360
gtaggccgac aggctcatgc cggccgccgc cgccttttcc tcaatcgctc ttcgttcgtc 420
tggaaggcag tacaccttga taggtgggct gcccttcctg gttggcttgg tttcatcagc 480
catccgcttg ccctcatctg ttacgccggc ggtagccggc cagcctcgca gagcaggatt 540
cccgttgagc accgccaggt gcgaataagg gacagtgaag aaggaacacc cgctcgcggg 600
tgggcctact tcacctatcc tgcccggctg acgccgttgg atacaccaag gaaagtctac 660
acgaaccctt tggcaaaatc ctgtatatcg tgcgaaaaag gatggatata ccgaaaaaat 720
cgctataatg accccgaagc agggttatgc agcggaaaag atccgtcgga tct          773


<210> 6
<211> 2012
<212> DNA
```

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Psi C31
      integrase.

<400> 6
```
agatctcccg tactgacgga cacaccgaag ccccggcggc aaccctcagc ggatgccccg   60
gggcttcacg ttttcccagg tcagaagcgg ttttcgggag tagtgcccca actggggtaa  120
cctttgagtt ctctcagttg ggggcgtagg gtcgccgaca tgacacaagg ggttgtgacc  180
ggggtggaca cgtacgcggg tgcttacgac cgtcagtcgc gcgagcgcga gaattcgagc  240
gcagcaagcc cagcgacaca gcgtagcgcc aacgaagaca aggcggccga ccttcagcgc  300
gaagtcgagc gcgacggggg ccggttcagg ttcgtcgggc atttcagcga agcgccgggc  360
acgtcggcgt tcgggacggc ggagcgcccg gagttcgaac gcatcctgaa cgaatgccgc  420
gccgggcggc tcaacatgat cattgtctat gacgtgtcgc gcttctcgcg cctgaaggtc  480
atggacgcga ttccgattgt ctcggaattg ctcgccctgg gcgtgacgat tgtttccact  540
caggaaggcg tcttccggca gggaaacgtc atggacctga ttcacctgat tatgcggctc  600
gacgcgtcgc acaaagaatc ttcgctgaag tcggcgaaga ttctcgacac gaagaacctt  660
cagcgcgaat tgggcgggta cgtcggcggg aaggcgcctt acggcttcga gcttgtttcg  720
gagacgaagg agatcacgcg caacggccga atggtcaatg tcgtcatcaa caagcttgcg  780
cactcgacca ctcccttac cggacccttc gagttcgagc ccgacgtaat ccggtggtgg  840
tggcgtgaga tcaagacgca caacacctt cccttcaagc cgggcagtca agccgccatt  900
cacccgggca gcatcacggg gctttgtaag cgcatggacg ctgacgccgt gccgacccgg  960
ggcgagacga ttgggaagaa gaccgcttca agcgcctggg acccggcaac cgttatgcga 1020
atccttcggg acccgcgtat tgcgggcttc gccgctgagg tgatctacaa gaagaagccg 1080
gacggcacgc cgaccacgaa gattgagggt taccgcattc agcgcgaccc gatcacgctc 1140
cggccggtcg agcttgattg cggaccgatc atcgagcccg ctgagtggta tgagcttcag 1200
gcgtggttgg acggcagggg gcgcggcaag gggctttccc ggggggcaagc cattctgtcc 1260
gccatggaca agctgtactg cgagtgtggc gccgtcatga cttcgaagcg cggggaagaa 1320
tcgatcaagg actcttaccg ctgccgtcgc cggaaggtgg tcgacccgtc cgcacctggg 1380
cagcacgaag gcacgtgcaa cgtcagcatg gcggcactcg acaagttcgt tgcggaacgc 1440
atcttcaaca agatcaggca cgccgaaggc gacgaagaga cgttggcgct tctgtgggaa 1500
gccgcccgac gcttcggcaa gctcactgag gcgcctgaga agagcggcga acgggcgaac 1560
cttgttgcgg agcgcgccga cgccctgaac gcccttgaag agctgtacga agaccgcgcg 1620
gcaggcgcgt acgacggacc cgttggcagg aagcacttcc ggaagcaaca ggcagcgctg 1680
acgctccggc agcaaggggc ggaagagcgg cttgccgaac ttgaagccgc cgaagccccg 1740
aagcttcccc ttgaccaatg gttccccgaa gacgccgacg ctgacccgac cggccctaag 1800
tcgtggtggg ggcgcgcgtc agtagacgac aagcgcgtgt cgtcgggct cttcgtagac 1860
aagatcgttg tcacgaagtc gactacgggc aggggcagg gaacgcccat cgagaagcgc 1920
gcttcgatca cgtgggcgaa gccgccgacc gacgacgacg aagacgacgc ccaggacggc 1980
acggaagacg tagcggcgta gcgagacacc cg                               2012
```

<210> 7
<211> 84
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Left arm of
       pPAOI6 transposon.

<400> 7
ctgtctctta tacacatctc aaccatcatc gatgaatttt ctcgggtgtt ctcgcatatt 60
                                                                      84
ggctcgaatt cgagctcggt accc

<210> 8
<211> 94
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Rigth arm of
       pPAOI6 transposon.

<400> 8
gatcctctag agtcgacctg caggcatgca agcttgccaa cgactacgca ctagccaaca 60
                                                                      94
agagcttcag ggttgagatg tgtataagag acag

<210> 9
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer AmF.

<400> 9   31
ccctaagatc tggttcatgt gcagctccat c

<210> 10
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer AmR.

<400> 10   34
tagtacccgg ggatccaacg tcatctcgtt ctcc

<210> 11

```
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer OriTF.

<400> 11                                                    32
gcggtagatc tgtgatgtac ttcaccagct cc


<210> 12
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer OriTR.

<400> 12                                                    37
tagtacccgg ggatccgacg gatcttttcc gctgcat


<210> 13
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer Fint.

<400> 13                                                    32
aacaaagatc tcccgtactg acggacacac cg


<210> 14
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer PJ.

<400> 14                                                    21
cgggtgtctc gcatcgccgc t


<210> 15
```

<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
      amyE-BamHI

<400> 15   30
atcgcaggat cctgaggact ctcgaacccg

<210> 16
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
      amyE-EcoRI

<400> 16   37
cgactgaatt cagatctagc gtgtaaattc cgtctgc

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   C12N15/10   C07K14/00   C07H21/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   C12N   C07K   C07H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, EPO-Internal, BIOSIS, MEDLINE, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 78977 A (AUGUST PAUL ;MACNEIL IAN (US); GROSSMAN TRUDY (US); AVENTIS PHARMA) 28 December 2000 (2000-12-28) page 5, line 1 – line 11 page 11, line 17 – line 19 page 14, line 13 – line 17 claim 21 ——— -/--	1-5,7,8, 18,19

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 January 2004	23/01/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Voigt, H

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	MEJIA J E ET AL: "RETROFITTING VECTORS FOR ESCHERICHIA COLI-BASED ARTIFICIAL CHROMOSOMES (PACS AND BACS) WITH MARKERS FOR TRANSFECTION STUDIES" GENOME RESEARCH, COLD SPRING HARBOR LABORATORY PRESS, US, vol. 7, no. 2, February 1997 (1997-02), pages 179-186, XP001106025 ISSN: 1088-9051 page 179, column 2, line 14 - line 21 page 181, column 2, paragraph 2 page 183, column 2, paragraph 1 page 184, column 2, line 20 - line 27	1,2,4,5
X	KIM SU YOUNG ET AL: "Modification of bacterial artificial chromosome clones using Cre recombinase: Introduction of selectable markers for expression in eukaryotic cells." GENOME RESEARCH, vol. 8, no. 4, April 1998 (1998-04), pages 404-412, XP002232358 ISSN: 1088-9051 page 409, column 1, line 3 - line 8	1,4-6
X	DATABASE EMBL 'Online! 23 November 2000 (2000-11-23) "Arabidopsis thalania chromosom 1" retrieved from EBI Database accession no. AC084820 XP002232359 abstract	22
Y	EP 0 176 321 A (LILLY CO ELI) 2 April 1986 (1986-04-02) page 3, line 23 - line 28	1-21
Y	WO 98 17811 A (CHROMAXOME CORP) 30 April 1998 (1998-04-30) page 7, line 28 - line 33 page 9, line 8 - line 11	1-21
A	WO 01 40497 A (AVENTIS PHARMA SA ;PERNODET JEAN LUC (FR); COURTOIS SOPHIE (FR); J) 7 June 2001 (2001-06-07) the whole document	1-21

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	RONDON M R ET AL: "CLONING THE SOIL METAGENOME: A STRATEGY FOR ACCESSING THE GENETIC AND FUNCTIONAL DIVERSITY OF UNCULTURED MICROORGANISMS" APPLIED AND ENVIRONMENTAL MICROBIOLOGY, WASHINGTON,DC, US, vol. 66, no. 6, June 2000 (2000-06), pages 2541-2547, XP000946829 ISSN: 0099-2240 the whole document —————	1-21

FURTHER INFORMATION CONTINUED FROM   PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-21

> Method of analysing a library of polynucleotides comprising the selection of cloning vectors, modification of these vectros and analysis the polynucleotides contained in vectors upon transfer in host cell.

2. Claims: 22-23

> Polynucleotide sequence comprising all or part of SEQ ID NO:1 or 2. An oligonucleotide comprising SEQ ID NO:3 or 4.

INTERNATIONAL SEARCH REPORT

Box I   Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II   Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

```
see additional sheet
```

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest   ☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (1)) (July 1998)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0078977	A	28-12-2000	AU	5496600 A	09-01-2001
			BR	0011765 A	05-03-2002
			CA	2375821 A1	28-12-2000
			EP	1196613 A1	17-04-2002
			JP	2003503029 T	28-01-2003
			NO	20016153 A	15-02-2002
			WO	0078977 A1	28-12-2000
			ZA	200109441 A	17-02-2003
EP 0176321	A	02-04-1986	AU	592956 B2	01-02-1990
			AU	4787785 A	10-04-1986
			CA	1238591 A1	28-06-1988
			DE	3572510 D1	28-09-1989
			DK	429685 A	28-03-1986
			EP	0176321 A1	02-04-1986
			HU	38676 A2	30-06-1986
			IE	58265 B1	25-08-1993
			JP	8015432 B	21-02-1996
			JP	61085191 A	30-04-1986
			US	4921801 A	01-05-1990
WO 9817811	A	30-04-1998	US	5783431 A	21-07-1998
			AU	744960 B2	07-03-2002
			AU	5163298 A	15-05-1998
			CA	2269123 A1	30-04-1998
			EP	0951557 A1	27-10-1999
			JP	2001502547 T	27-02-2001
			KR	2000052795 A	25-08-2000
			WO	9817811 A1	30-04-1998
			US	6242211 B1	05-06-2001
			ZA	9709557 A	12-05-1998
WO 0140497	A	07-06-2001	FR	2801609 A1	01-06-2001
			AU	2179101 A	12-06-2001
			BR	0015993 A	06-08-2002
			CA	2393041 A1	07-06-2001
			EP	1268764 A2	02-01-2003
			WO	0140497 A2	07-06-2001
			JP	2003520578 T	08-07-2003
			NO	20022532 A	26-06-2002